DOCUMENT RESUME

ED 189 819 EC 124 550

AUTHOR MacIntyre, Robert B.; And Others
TITLE Identification of Learning Disabilities in Ontario: A Validity Study.
INSTITUTION Ontario Dept. of Education, Toronto.
REPORT NO 371.26013
PUB DATE 80
NOTE 146p.
AVAILABLE FROM Publication Sales, The Ontario Institute for Studies in Education, 252 Bloor Street W., Toronto, Ontario, Canada M5S 1V6; The Ontario Government Bookstore, 880 Bay St., Toronto, Ontario, Canada M7A 1L2 ($3.00)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Achievement Tests; *Definitions; Elementary Secondary Education; Evaluation Methods; Exceptional Child Research; Foreign Countries; *Handicap Identification; Intelligence Tests; *Learning Disabilities; Perception Tests; Psycholinguistics; *Screening Tests; *Test Validity
IDENTIFIERS Ontario (Toronto)

ABSTRACT

The study examined the process by which diagnostic data on children thought to have learning disabilities (LD) are generated and used. Emphasis was placed on the validity of test instruments in common use in the field. Since part of the practical validity of any instrument used to identify an LD child rests on the validity of the LD construct itself, current definitions of LD were examined for their implications for evaluative measures. Fifteen tests were reviewed: four intelligence measures (Wechsler Intelligence Scale for Children-Revised, Denver Developmental Screening Test, Raven's Coloured Progressive Matrices, Slosson Intelligence Test); two tests of psycholinguistic abilities (Illinois Test of Psycholinguistic Abilities, McCarthy Scales of Children's Abilities); five visual and auditory perception tests (Bender-Gestalt Test, Slingerland Screening Test for Identifying Children with Specific Language Disability, Frostig Developmental Test of Visual Perception, Beery-Buktenica Developmental Test of Visual Motor Integration, Wepman Auditory Discrimination Test); and four standardized achievement tests (Wide Range Achievement Test, Canadian Test of Basic Skills, Gates-MacGinitie Reading Tests, Peabody Individual Achievement Test). Among common components of definitions in use were a disparity between some measure of potential and some measure of performance and an assumption of an underlying dysfunctonal learning process. The majority of tests in use had not been normed on an LD population, and no single test used alone was found to be a valid assessment tool. Visits to selected Ontario School Boards revealed that definitions and assessment procedures were not uniform across the province. (SBH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189819

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IDENTIFICATION OF LEARNING DISABILITIES IN ONTARIO

A Validity Study

ROBERT B. MacINTYRE, Principal Investigator
ANNE KEETON
RALPH AGARD

This study reflects the views of the authors and not necessarily those of the Ministry of Education.

This research project was funded under contract by the Ministry of Education, Ontario.

Hon. Bette Stephenson, M.D., Minister
Dr. H. K. Fisher, Deputy Minister

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Minister of
Education.
Ontario
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EC124550

2

JUN 2 1980

© The Minister of Education, Ontario, 1980
Queen's Park
Toronto, Ontario

MINISTRY OF EDUCATION, CATALOGUING IN PUBLICATION
MacIntyre, Robert B.
Identification of learning disabilities in Ontario : a validity study / Robert B. MacIntyre, principal investigator; Anne Keeton ; Ralph Agard.--

This research project was funded under contract by the Ministry of Education, Ontario.

1. Educational tests and measurements - Ontario. 2. Psychological tests for children - Ontario. 3. Learning disabilities. I. Keeton, Anne. II. Agard, Ralph. III. Ontario. Ministry of Education. IV. Title.

ONO 1675 371.26013

ISBN: 0-7743-4901-8

Additional copies may be ordered from:

Publications Sales
The Ontario Institute for Studies in Education
252 Bloor Street West
Toronto, Ontario
M5S 1V6

or from:

The Ontario Government Bookstore
880 Bay Street
Toronto, Ontario
M7A 1L2

1 2 3 4 5 XX 48 38 28 18 08

Abstract

The purpose of this study was to examine the process by which diagnostic data on children thought to have learning disabilities is generated and used. Emphasis was placed on the validity of test instruments in common use in the field. A small sample of Boards were visited to study the learning disability identification and placement procedure.

Since part of the practical validity of any instrument used to identify a child with learning disabilities rests on the validity of the learning disability construct itself, current definitions of learning disability were examined for their implications for evaluative measures. Fifteen tests, representing those most in use in the field, were reviewed: four intelligence measures, two measures of psycholinguistic abilities, five tests of auditory and visual perception, and four standardized achievement tests.

Three major components were found to be common among all the definitions in use, including that of the Ministry of Education. These were: a disparity between some measure of potential and some measure of performance; an assumption of an underlying dysfunctional learning process; and the exclusion of children whose learning pattern could be ascribed to other conditions. In addition, the definitions in use were found to emphasize present functioning and, in some cases, causes or etiology. None contained any components dealing with possible treatments or outcomes either with or without treatment. Existing definitions of the construct "learning disability" were found to be seriously flawed and to present major difficulties to the designers and users of assessment tools and tests.

The tests reviewed, to a greater or lesser extent, also were found to have serious flaws when used with a learning disability population. The majority of instruments in use had not been normed on a learning disability population, including the heavily used Illinois Test of Psycholinguistic Abilities. Where predictive validity studies were available for a learning disability sample, the results were not encouraging. No single test used alone was found to be a valid assessment tool. Validity reports for each test reviewed are included in the study.

Visits to selected Boards revealed that definitions in use and assessment procedures were not uniform across the Province. Most Boards used some form of two-level assessment: an inschool level, primarily educational in nature; and a second level, primarily psychological and based on standardized tests. At this second level, little use was made of teacher input and participation.

The study concludes that the identification and placement of children with learning disabilities in Ontario must continue to be based on informed judgement using as wide a variety of data bases as possible rather than rely on single indices or standardized test scores.

Contents

Acknowledgements i

Introduction 1

Procedures 5

Definitions 7

Universal Components of Learning Disability Definition 9

Learning Disabilities Definitional Commonalities and (Resulting) Important Implications for Evaluative Measures 21

Implications for Evaluative Measure 24

Individual Intelligence Tests 27

Wechsler Intelligence Scale for Children, Revised 29

Denver Developmental Screening Test 38

Coloured Progressive Matrices (Raven's) 43

Slosson Intelligence Test 49

Tests of Psycholinguistic Abilities 51

Illinois Test of Psycholinguistic Abilities 51

McCarthy Scales of Children's Abilities 59

Visual and Auditory Perception Tests 62

Bender-Gestalt Test 62

The Screening Test for Identifying Children with Specific Language Disability (Slingerland) 64

Developmental Test of Visual Perception (Frostig) 66

Developmental Test of Visual Motor Integration (Beery-Buktenica) 68

  Auditory Discrimination Test (Wepman) 70

  Conclusion 74

Standardized Achievement Tests 75

  Wide Range Achievement Test 76

  Canadian Test of Basic Skills 80

  Gates-MacGinitie Reading Tests 82

  Peabody Individual Achievement Test 86

Predictive Validity of Groups of Tests, including Standardized Achievement Tests, for Disabled Learners 89

The Learning Disabled Pupil in Ontario:

  Report of Interviews Conducted on a Sample of Ten Ontario School Boards 95

  Definitions, Terminology and Evaluative Criteria 101

  The Assessment Process 107

  Instruments Used 109

  Summary 111

Summary and Conclusions 114

  Definitional Programs in Learning Disabilities 114

  Validity of Assessment Procedures 116

  Toward an Alternative Look at Assessment 117

  Final Note 122

Index to Tests Reviewed 124

References 125

List of Tables

Table 1 Boards Identified with Segregated Learning Disability Classes and Persons who do Diagnostic Testing 97

Acknowledgements

The authors wish to thank the following for their contributions to this paper:

Marta Valenzuella and Dana McLaren, for work on the test validity issues;

Rosemary Grey-Snellgrove and Karen Dunbar, for typing and editing services;

And our students in Assessment courses at the Ontario Institute for Studies in Education, for helping us examine the issues in this complex field.

R.B.M.

Introduction

This project was undertaken to assess the validity of instruments currently in use in Ontario for the diagnosis, evaluation and placement of children with learning disabilities. Children identified as having learning disabilities now constitute a sizable and growing proportion of those receiving special education services in the Province of Ontario. The growth of this area has been marked by varied definitions and equally varied assessment procedures. In the absence of accepted guidelines, school boards have independently developed their own assessment systems, using a variety of instruments, administered by individuals with various types of training, and interpreted and applied by various types of committees. At present the University of Ottawa is engaged in a Learning Disabilities Research Project designed to yield information on the types of programs and services provided by school boards, including the identification of the categories of personnel involved in assessment. Another survey, by Anne Keeton, had examined the special education assessment process in general, including the learning disabilities area. From preliminary examination of these and other data, it was apparent that wide disparity exists among boards in their process of identifying, placing and programming for children with learning disabilities.

Two major aspects of this disparity, both relating to the validity of the process, formed the focus of the study. The first is the validity of the instruments involved when used with children with learning disabilities. A wide variety of tests are in use, drawn from diverse sources; some initially designed to identify adult brain injury, to measure intelligence in a normal population, to identify psychodynamic processes, to determine academic achievement, and a few even designed to guide educational programming for students with learning problems. The established validity of these instruments varies greatly when used for their initial purpose and even more so when used with children having learning disabilities. From our experiences in teaching courses in assessment over the last few years, we knew that many people in school boards were using tests without a clear understanding of the issue of validity and certainly without specific knowledge of the validity of a particular instrument for the purpose to which it has been put.

The second major validity issue had to do with the process by which diagnostic data is generated and used in schools. Presently most decision-making regarding learning disability programming involves some sort of team effort. Of necessity then, assessment data is read and interpreted by individuals other than the tester. Committees may attempt to justify decisions on the basis of test results which do not relate to that particular decision, and data collected for one purpose may be used for another. Some of the potential sources of invalidity in the process are:

1. Selection of an inappropriate instrument for a particular purpose.

2. Inappropriate administration of the test, including use of an English-language and middle-class-biased test exclusively when assessing children with varied language, cultural and economic backgrounds.

3. Inappropriate interpretation of test results by the initial assessor.

4. Inappropriate interpretation by other individuals on the team at a later point. This is particularly likely in regard to interpretation of test IQ scores by school personnel and parents lacking psychology training.

5. Inappropriate application of assessment data, as in using diagnostic instruments, finely tuned to show small variations from "normal", for selection purposes. Many successful university graduates can be identified as having a "learning disability" on some of the tests currently in use in our schools.

In looking at this assessment process, two major uses of assessment data were apparent. The first use was in identification of individuals with learning disabilities. Here we may have screening instruments, early identification instruments and other tools and processes by which we make the initial decision to provide special education services. At one time the assumption was that such processes should throw a wide net, capturing all individuals who might fit, at the risk of including others who might not - the false positives, as the testers say. This was in the belief that no one would be damaged by special educational help and that we should provide this help as soon as possible to as many as possible. Two problems have appeared with this approach. One is the problem of identification without programming. Children may receive no more than "waiting list treatment" if the assessment program outruns educational programming, or if the child is defined as having a problem for which no program is presently available. The second problem to emerge is that of the stigmatization of the child because of the assessment and categorization process. It is becoming evident that certain labels used in special education influence both the attitudes of teachers and others toward a child and the child's own motivation and self-concept. Consequently, a desirable assessment process must steer a middle course between the risk of missing children who need help and wrongly identifying others. Ideally, tests should be selected which have high discriminating ability and yield a low proportion of both false positives and false negatives, a difficult and sometimes impossible task.

In summary, the two aspects of validity which are considered in this study are the validity of the instruments used with learning disability children in Ontario, and the validity of the assessment systems by which these data are generated, interpreted and applied.

Procedures

Step one in the process of assessing the validity of instruments and systems in use in Ontario for the identification and programming of children with learning disabilities was to determine which instruments were presently being used and what background the testers had. The Learning Disability Research Project at the University of Ottawa had collected some data on the assessment process, as had Anne Keeton at O.I.S.E. Data from these sources, including copies of selected school board referral and placement forms, were examined. Since neither project had focused on tests in use, it was necessary to telephone a small sample of boards in order to develop the full list of tests to be evaluated.

Since many tests have been standardized and used for some time, validity studies had been done. A simple check of Buros Mental Measurement Yearbooks was sufficient to answer the validity questions involved, since learning disabilities is a fairly recent category of exceptionality and one not specifically considered by earlier test evaluators. Therefore a comprehensive search was made in the special education and testing literature for validity reports in the learning disability literature. ERIC and EISO were used in this search, and references in the recent Harvard Educational Review article (Coles, 1978) on this issue were examined. One hundred and twenty-two studies were discovered, of which 60 were directly relevant to the purposes of this study.

Since the second issue in this study is the process by which school boards use test data for decision-making about children with learning disabilities, a small selection of eleven school boards were visited and their procedures probed in some depth by interview. Boards were chosen on the basis of returns in both the Ottawa and O.I.S.E. studies to reflect various extremes in referral rates, choice of personnel and programming options. No attempt was made to choose or identify "best" or "ideal" processes; rather boards were selected whose assessment systems were likely to yield comparisons which would contribute to our understanding of common processes and variations. This phase of the study also examined the various interpretations of definitions in use for the learning disability area and the relationship between definition and assessment procedures.

Definitions

Within the past decade the field of learning disabilities has grown tremendously. Its growth has been accompanied by difficulties in arriving at a general consensus regarding definition, etiology, diagnostic procedures and measures, treatments, and prognosis.

The appropriate diagnosis of a learning disability hinges upon the specification of just what a learning disability may be. If this were clearly established, the accuracy of the measures used to ascertain its presence could be more easily analyzed.

The question, "What is a learning disability?", is not easily answered. A review of the literature shows that a consensus on a functional definition is non-existant. Recently, contributors to the field have demanded that steps be taken to clarify some of the ambiguities and contradictions which presently exist (Gomez, 1967; Orlando and Lynch, 1974).

It is perhaps wise, before embarking on a project of this type, to establish some ground rules or baselines which would ensure consistency in the evaluation of different measures.

For this purpose, the authors will briefly attempt to outline the type of definitions in use and abstract some commonalities. It is believed that isolation of these commonalities would delineate specific contingencies by which one may evaluate the measures used in the learning disabilities diagnostic procedure.

The National Advisory Committee on Handicapped Children (1968) outlined a definition of learning disabilities which can provide a branching point for consideration of other definitions. This definition reads:

> Children with special learning disabilities exhibit a disorder in one or more of the basic psychological processes involved in understanding or using spoken or written languages. These may be manifested in disorders of listening, thinking, talking, reading, writing, spelling, perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, developmental aphasia etc. They do not include learning problems which are due primarily to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to environmental disadvantage.

This definition is characterized by the exclusion approach. Coles (1978) exemplifies the feelings of many professionals in the field when he states, "It is apparent that this definition has its limitations. What a learning disability is remains extraordinarily vague; it is primarily a definition by exclusion."

In spite of its limitations, other learning disability experts adhere to definitions which closely follow the National Advisory Committee's definition. They have, however, very carefully added a conditional note preceding the definition to the effect that a distinction has to be made between learning problems in school and specific learning disabilities.

Kirk and Bateman (1962) preface their definition of a learning disability by stating that "while there are a substantial number of children who are delayed or retarded in learning to talk, read, write, spell, or do arithmetic, not all children with these problems are considered to have learning disabilities per se." Johnson and Morasky (1977) argue that "a most common error is to assume that learning disabilities and learning problems encountered in everyday school experiences are identical", while Myers and Hammill (1969) indicate that a "learning disability is used to describe a specific type of exceptional child, it is not a generic term for all children who have learning problems in school."

Kirk and Bateman, in addition to their prerequisite which attempts to distinguish between a specific learning disability and some problems encountered in everyday school experiences, differ from the National Advisory Committee's definition by including an etiological factor. Inherent in their definition is the fact that the delayed development in learning abilities is the result of "a psychological handicap caused by a possible cerebral dysfunction and/or emotional or behavioral disturbance." They, however, retain the exclusion of etiological factors of mental retardation, sensory deprivation, and cultural or instructional factors.

Universal components of learning disability definition. The U.S. Office of Education ordered a committee to write a definition which would be more universally acceptable.

The following points resulted from this conference (Kass & Myklebust, 1969):

> Learning disability refers to one or more significant deficits in essential learning processes requiring special education for remediation.
>
> Children with learning disability generally demonstrate a discrepancy between expected and actual achievement in one or more areas such as spoken, read, or written language, mathematics, or spatial orientation.
>
> The learning disability referred to is not primarily the result of sensory, motor, intellectual, or emotional handicap, or lack of opportunity to learn.
>
> Significant deficits are defined in terms of accepted diagnostic procedures in education and psychology.
>
> Essential learning processes are those currently referred to in behavioral sciences as involving perception, integration, and expression of language and meaning, either verbal or nonverbal.
>
> Special education techniques for remediation refers to educational planning based on the diagnostic procedure and results. (Kass & Myklebust, 1969)

Of particular interest in these universal definitional components is the term "accepted diagnostic procedures in education and psychology." Hammill (1972) points out that it is very difficult to gain a consensus on the terms "psychological process" or "perceptual functioning." Larsen et al. (1976), referring to both the terms "psychological process" and "perceptual functioning," state, "these frequently used terms have a tendency to be all-encompassing. It would be possible to say that all children with a learning difficulty have processing and/or perceptual deficiencies. To do so however negates the meaning of the terms and limits their use in a definition" (Larsen, 1976).

Another criticism of the exclusion definition is its role as an additive force in the labelling process (for convenience reasons), and the possible elimination of the poor and culturally different from appropriate learning disability programming. On the exclusion principle, Coles (1978) suggests that "the specific conditions mentioned, such as dyslexia, are no more precisely defined than the encompassing term 'learning disabilities', and by excluding environmental disadvantage it (the definition) limits attention to middle and upper class children."

The exclusion principle however, is in reality inclusive in function. It may in fact be an inclusion principle. The present special education format, borrowed from the medical model, is based upon disability categories. In order for a student with special needs to receive any type of educational attention, he must be identified and labelled. However, as Blatt (1972) argues, labelling "dehumanizes and stigmatizes both these children and their families" (p.542).

Mercer (1971) points out that the labelling approach has resulted in the frequent placement of children of minority groupings into one of these exceptional categories.

Some educators believe that exclusion definitions are in fact attempts at locating ways of including, for purposes of additional programming, children from middle and high middle class backgrounds, who have failed to achieve academic success in the present educational system (Farrald & Schamber, 1973; Valett, 1969).

Thus, the underlying assumption is that by defining these students as learning disabled, and having them identified, adequate programming will follow. This concept is further evidenced by the writings of Foster et al. (1976). "Some authors suggest that the labelling of a child as 'learning disabled' is somehow preferable to other disability groupings." He further points out Ford's (1971) contention that the definition of learning disability "implies the presence of normal intelligence", and that Lerner (1971) reported that "parents of children with academic problems have welcomed the advent of the learning disabilities movement and have been largely responsible for its impetus." Foster and Sabatino (1976) attest "that parents often accept the label of learning disabled as preferable to other categories of disability, presumably because it is a less stigmatizing label, stressing normal or above normal intelligence" (p. 59).

Hallahan and Cruickshank (1973) state the case for the denial of the poor and culturally different from programming most eloquently. They state, "unfortunately, current definitions of learning disabilities exclude children with learning difficulties due to environmental influences. While they ensure that poor children will not be misplaced in classes for the learning disabled, they ignore those economically disadvantaged children who display behaviors identical to those of middle class children appropriately placed in learning disability classes" (p.13). They conclude further that "the consequence of all this is that many children of culturally different backgrounds are assigned to classes for the mentally retarded, while middle class children exhibiting the same behavioral characteristics are placed in classes for the learning disabled. It is, thus, our opinion that children who have learning problems due to environmental conditions should not be excluded from learning disability programs" (p.13).

From this brief review it can be seen that the definition of learning disability is as much social and political as it is clinical. The assumptions built into the exclusionary clauses, that either "below average intelligence" or "cultural deprivation" are categories or labels which take precedence over "learning disabilities" works to exclude the former groups from the latter and suggests that one can have only one of these problems. In the extreme this type of definition would inevitably produce "learning disability" classes composed only of children of average intelligence from middle-class families who are not doing well in school. The emphasis on defining a clinical entity without proper attention to the defining process and the role of the definers (Mercer, 1971) is a serious and basic flaw in the major existing definitions.

Other educators and researchers place a heavy emphasis on the neurological etiology of a learning disability (Clements, 1966; Cruickshank, 1973; Johnson & Myklebust, 1967; Mercer, Forgone & Wolking, 1976).

In relation to the role of the central nervous system in the concept of a perceptual deficit (with or without appropriate evidence), this is a predominant feature of many learning disability definitions, either by inclusion or by assertion. The work of Strauss (1955), Werner (1942), Cruickshank (1973), Kephart (1971), Frostig (1966), Ayres (1972) and others demonstrates the amount of energy which has been invested in the concept of a direct link between central nervous system disorder and perceptual dysfunction.

Cruickshank (1973) provides a distinct view of this concept. He states, "Thus, three major aspects of human development, either separately or in combination with one another, may be influential in producing neurological damage and a resulting perceptual dysfunction related to learning disorders. These include (1) specific neurological issues, (2) nutritional deprivation, and (3) factors of environmental deprivation." This statement satisfies two needs at this point; firstly it provides an example of the type of thinking consistent with the neurological-perceptual connection and secondly an overall review of the definitional constructs which we have considered thus far.

In addition to these already identified components, another descriptive construct, and perhaps the one most frequently used in current definitions, is the discrepancy concept. Two assertions emerge from this approach. One is the idea that a learning disabled child must be of average or above average intelligence. The second implies that a discrepancy between potential and achievement demonstrates the presence of a learning disability.

Two examples of definitions by McIntosh and Dunn (1973) and Kirk (1972) illustrate the second of these aforementioned assertions. McIntosh and Dunn (1973) suggest that children who exhibit major specific learning disabilities are those "who display a discrepancy between capacity and performance in a learning process, associated with the areas of oral and written language, or mathematics." Kirk (1972) contends that a child with a learning disability shows "developmental discrepancies in ability, has a specific problem that is not a correlate of other primary handicapping conditions, and displays behavioral deficits."

Bateman's (1964) definition for learning disabled children alludes to the fact that for these children there exists a "significant discrepancy between their estimated intellectual potential and actual level of performance related to basic disorders of the learning process." This concept is also supported by Gallagher (1966), Kass (1969), Kirk (1962) and Gearheart (1973). Gearheart (1973) describes a learning disabled child as "having average or above average intelligence, adequate sensory acuity...achieving considerably less than a composite of his intelligence, age, and education ability would predict."

In contrast to all the definitions previously considered, there is an alternative form which might be termed a "practical educational approach". Proponents of this approach question the relationship of etiology and the development and delivery of appropriate remedial programs. The problems, they maintain, demonstrate the same observable behavior.

Hallahan and Cruickshank (1973) summarize this concept when they say, "reading problems, emotional problems, intellectual problems, speech problems, handwriting problems, and others, irrespective of their etiology or symptomatology, are found grouped together on the premise that each is a learning problem."

Valett takes the disparity approach and superimposes it upon the practical educational approach. He contends:

> A learning disability is here defined as any specific difficulty in acquiring and using information or skills that are essential to problem solving. A significant learning disability exists when the individual's actual performance or achievement in any given area is found to be far below his capacity

> or potential. A discrepancy between classroom achievement and mental age of one or more years is usually cause for concern, while a discrepancy of two or more years calls for immediate educational intervention. (p. 3)

Kass and Myklebust express the same opinion, but take an even heavier practical educational approach when they describe a learning disability as "one or more significant deficits in essential learning processes requiring special education techniques for their remediation. Children with a learning disability generally demonstrate a discrepancy between expected and actual performance in one or more areas, such as spoken, reading or written language, mathematics, and spatial orientation" (Kass & Myklebust, 1969).

A summary of this section on definitions must demonstrate two basic concepts. Firstly, to date there is no consensus on a definition of learning disabilities. Secondly, the areas of greatest concern among experts which lead to different orientations of definitions are:

1. The exclusion of culturally disadvantaged children and environmental etiological factors may defeat the original intent of the exclusion principle.

2. The exclusion principle may in fact be considered an inclusion principle, since in reality it allows middle- and upper middle-class students to receive remedial programming, who prior to this were unable to do so.

3. The terms "psychological dysfunction" and "perceptual

dysfunction" are too vague, and need greater precision if they are to be used in definitions.

4. The etiological factor approach appears to conflict with the practical educational approach. Proponants of either school of thought attempt to demonstrate the need for a heavy emphasis on their individual orientation.

Some clarity on this issue may be obtained by using the analysis provided by Cromwell et al. (1975). They analyse the elements of a diagnostic construct as follows: a statement or knowledge of the etiology or causative agent of the condition (A); a statement of the present level of functioning of the individual, including present symptomatology (B); a statement of the treatment or remediation procedures to be used to ameliorate the condition (C); and (D), a statement of the prognosis or predicted outcome of the condition. In their view any diagnostic condition which does not include either C or D is invalid and not useful in any practical sense. Thus we may have a condition where we are ignorant of the cause but do know the symptoms and treatment, or prognosis, as in the case of smallpox before the discovery of viruses. We may also have conditions for which no treatment exists but where we can make clear predictive statements of the outcome, as in spinal chord injury. Either situation provides a valid diagnostic label or category.

Problems exist, however, when our label or category includes only A and B with no clear link to either treatment or prognosis. We would (and do) think poorly of a physician who tells us that we have "X" condition because of "Y"; that he is not sure that any treatment exists and that we may or may not recover anyway.

By focusing our definitions on only statements of present functioning, e.g. discrepancy between expected and actual school performance or performance on perceptual motor or psycholinguistic tests; and on presumed etiological constructs, e.g. neurological integration, brain dysfunction, etc., we have constructed most of our definitions in an invalid form in Cromwell's (1975) sense. The literature on the links between specific psychological, perceptual, and psychological processes; the many and varied intervention programs designed to treat or remedy them; and the effects on school learning are outside the scope of this study. We can say that there is little consistent evidence that any clear link has been found between our present treatment procedures, any specific etiology and learning outcomes. This is not to say that our special education programs have not been effective because they have helped many children. It is to say that the improvements noted have not been specific to the conditions defined. To pursue the medical analogy, we seem to be in the stage of having some moderately effective antibiotic around, which cures a variety of conditions, and of being tempted to use it for "general prophylaxis".

Thus, even today as the level of sophistication in the field continues to grow, bulwarked by more research and organized study, there is still a great deal of confusion and disagreement over the actual boundaries which limit the field and the lack of a single, universally accepted definition. Discouraging as this may seem at times, it is our opinion that effective remediation and productive research are possible nevertheless. Those choosing to work in the field should orient themselves toward immediately demonstrable and defensible methods. They may therefore actually benefit from the diversity of opinion through carefully planned and objective eclecticism while remaining sensitive to the possible limitation as well as the advantage of different ideas and approaches (Johnson & Morasky, 1977).

An evaluatory diagnostic process is the core of any functionally effective remediation effort. If, as Johnson and Morasky (1977) suggest, workers in the field of learning disabilities must orient themselves toward immediately demonstrable and defensible methods, then the tests used in the learning disability diagnostic process must be subject to careful scrutiny. They must be evaluated for accuracy of pinpointing the deficits identified in learning disability definitions. Once this is accomplished, workers in the field may be able to select at least one demonstrable and defensible method.

Learning Disabilities Definitional Commonalities and (Resulting) Important Implications for Evaluative Measures

Commonalities. Basic to the development of the field of learning disabilities is the belief that there exists a homogenous population of children who display some common characteristic features. This is accompanied by the belief that children who are learning disabled are identifiable and, once identified, their conditions are amenable to remediation.

The definitions used in the field are in fact statements of evidence of belief in such a premise. Despite the diversity of definitions in current usage, there do exist some commonly identified elements of learning disabilities and characteristics of learning disabled children. Several authors have supported this notion (Gearheart, 1973; Kirk, 1972; Myers and Hammill, 1969). These ideas may be summarized as follows.

1. The principle of disparity.

A brief synopsis of this principle is that for the learning disabled, there exists a significant discrepancy between performance and predicted potential. This 'performance' is expected as normalcy in the areas of (a) speech and language, (b) perceptual functioning (includes visual perception, perceptual motor, etc.), and (c) academic achievement. Deficits in all or one of these areas may constitute the phenotypic manifestations of learning disabilities.

2. The exclusion principle.

Most definitions do exclude some category of learner. Concern is expressed regarding who exactly should be excluded, with many authors pleading a strong, almost mandatory case for the inclusion of the culturally disadvantaged.

3. The dysfunctional learning process principle.

Common to all definitions is the conviction that the dysfunction, regardless of the etiology, is related to and is outwardly manifested by a retardation in or termination of a process involved in learning and must be relevant to educational accomplishment.

4. The general role of central nervous system dysfunction.

While the role of the central nervous system as an etiological factor is not denied by any of the investigators, the subsequent use of this information, and the design of remedial instruction based upon it is the aspect of this definitional component which is most debated.

The definition adopted by the Ministry of Education includes the concepts of a disparity between performance and ability and the exclusion principle as well as the dysfunctional learning process principle. It does drop the notion of central nervous system dysfunction:

> The Ministry of Education defines learning disabilities as: disorders in one or more of the processes involved in understanding or using symbols or spoken language. The disorders result in a significant discrepancy between academic achievement and assessed intellectual

> ability, with deficits in at least one of the following areas:
>
> - receptive language (i.e. listening, reading);
> - language processing (i.e. thinking, conceptualising, integrating);
> - expressive language (i.e. talking, spelling, writing);
>
>   and
> - mathematical computations.
>
> Such deficits become evident in both academic and social situations.
>
> The definition does not include children who have learning problems which are primarily the result of impairment of vision or hearing; motor handicaps; mental retardation; primary emotional disturbance; or environmental, cultural or economic disadvantage. (Ministry of Education, 1978, p. 2)

This definition places some very precise obligations on the evaluator or assessor, if he is to pursue his endeavors with vigor. These obligations are exemplified by a thorough and extensive search for diagnostic capabilities and precision of information provided in direct relation to the components included in the working definition. Both the diagnostic capabilities and precision of information, which tests used in the diagnosis of learning disabilities must provide, are identifiable by a consideration of the implications of components of the definition.

Implications for Evaluative Measure. Inherent in the disparity component is the concept of a discrepancy between achievement and potential. If we accept the disparity approach, then tests must not only demonstrate validity in assessing present conditions (achievement), but must also possess acceptable predictive validity (potential). This is necessary if one is to ascertain a discrepancy between potential and performance. More importantly, however, these characteristics should be demonstrated with particular reference to the learning disabled population.

Coles (1978) and Noonan (1977) identify the common tests used frequently in estimating the intellectual component. These tests are: The Wechsler Pre-School and Primary Scale of Intelligence, The Wechsler Intelligence Scale for Children, The Stanford-Binet, The Illinois Test of Psycholinguistic Abilities, Slossen, and The Picture Peabody Vocabulary Tests. When one takes cultural etiological factors into consideration, as has been recommended by some prestigious contributors to the field (Coles, 1978; Hallahan & Cruickshank, 1973), the cultural aspects of the intelligence test become important. The researcher must therefore consciously attempt to use tests for which the population upon which the tests were standardized is appropriate to the children being tested.

If intelligence tests are to be used as the signposts of potential, then they must maintain predictive validity, and the precise area of predictive competency established.

If the disparity is to be established in an area other than intellectual versus academic functioning, then these tests must establish predictive validity for those areas. In the event that such predictability is not established, the researcher is obligated to identify any correlations for performance in the specified area and another area for which accurate predictive measures exist. Because the definition includes an intelligence component, the most valuable studies to consider would be studies where the relationship between intelligence and achievement in children with learning disabilities was reported. The results from studies maintaining this type of experimental design would appear to be directly relevant to a learning disability population. Perceptual tests need to be considered for evidence of correlations between some of the subtests and academic achievement. They must also be assessed for their usefulness as an accurate predictor of reading or mathematical achievement.

A learning disability test battery is assumed to be not only diagnostic in nature, but also discriminatory. This implies that the tests assessed should demonstrate significant differences between learning disabled and normal children. This is particularly true for tests used to ascertain the domains of language and perceptual functioning. For these reasons, the variables controlled in establishing the relevant data must be taken into consideration. The exclusionary nature of many learning disability definitions makes it mandatory for the variables of intelligence, age, social class, and race to be controlled if accurate conclusions may be drawn which have relevance to the learning disabled.

Individual Intelligence Tests

Individual intelligence scales provide a major criterion for the identification of children with learning disabilities and are included in almost all test batteries in use in schools. Sometimes the IQ tests is used in conjunction with an exclusionary definition, to rule out the possibility that low or below normal intelligence can account for the learning problem. Often test indices and subtest scatter are used to indicate learning disability or retarded reader profiles. Sattler (1974) devotes 11 pages in his basic text on assessment to reporting the attempts to develop IQ test patterns indicative of learning disabilities and reading disabilities, criticizing the research for the lack of precision in identifying or defining the population of interest. His recommendations for extensive testing with a battery, including "the WISC (at least 10 of the subtests), Bender-Gestalt, Draw-a-Man, Gray Oral Reading Tests, and Wide Range Achievement Test" (Sattler, 1974, p. 332), have probably influenced the testing field significantly, but are based primarily on his observation that IQ measures alone do not adequately identify learning problems.

Part of the problem in using IQ tests as a diagnostic tool in identifying learning disabilities rests on the fact that all such tests were initially designed to measure, to some extent, a general factor of intelligence and all have been validated, somewhat incestuously, with one another. Both of these elements contribute to a low degree of specificity in what a particular item or subtest actually taps. As an example, Digit Span on the WISC-R is frequently cited as a subtest which is low in those with learning problems, yet it is not normally included in calculation of an IQ score because of its low correlation with other subtests. The sensitivity of this subtest to difficulties in immediate memory, to anxiety and to other problems, makes it more useful as a diagnostic indicator and less useful as an IQ measure.

A second difficulty in the use of IQ measures in the identification of learning disabilities has been the attempt to transfer and validate findings from the study of brain injury in mature individuals or of diagnosed organic brain damage in children. Although the Ontario definition of learning disabilities does not include a presumption of brain dysfunction, many of the recommended diagnostic criteria for learning disabilities have been drawn from this work.

Title of Test: Wechsler Intelligence Scale for Children (WISC-R) Revised 1974 edition

Author: David Wechsler

Publisher: The Psychological Corporation, New York

Purpose: To measure intelligence in children.

Type of test: Individual administration.

Population for Which Designed: Children from 6 to 16 years of age.

Time Required to Administer: 1 to 1.5 hours.

Nature of Content: The test consists of twelve subtests distributed in two scales, Verbal and Performance (six on the Verbal Scale, six on the Performance Scale). Only 10 of the 12 subtests, 5 verbal and 5 performance, must be administered in order to calculate a child's IQ score. Verbal and Performance subtests form scales from which a Verbal and a Performance IQ are obtained. The Full Scale, or Total IQ score, is calculated from the two subscale scores.

Practical evaluation: test materials are attractive, easy to manipulate and appropriate for examinees. The test Manual can be considered as a good source of information for its users, providing clear and thorough explanations on administration, scoring and interpretation of results and norm tables. The WISC-R Manual and record form represent an improvement over the early edition WISC, providing more explicit and comprehensive instructions for starting points, discontinuance points, scoring criteria and examples. The examiner is required to be highly trained in administering the test, and skilled in making evaluative judgement on the examinee's responses.

Norms: The Manual presents information on the standardization of the WISC-R. The norms are derived from a sample of 2,200 children representative of the U.S. population of children with respect to age, sex, race, father's occupation, and geographical area. The standardization sample represents an improvement over the WISC standardization sample by including, in proportion to their numbers in the population, non-white and bilingual children. However, children were only included providing they could speak and understand English.

Reliability: The Manual reports information on split-half reliability, stability and error measurement. Split half reliability coefficients, which provide a measure of internal consistency, were computed for the WISC-R individual subtests (except for Coding and Digit Span). The reliabilities were satisfactory to excellent, with the average coefficient ranging from .77 to .86 for the Verbal tests and from .70 to .85 for the Performance tests. Reliability coefficients for the Verbal, Performance, and Full Scale are reported across the entire age range of the sample, the average coefficients being .94, .99, and .96 respectively. Stability coefficients, or the test's stability over a one-month period, were assessed for the three scales. The coefficients were .93, .90, and .95.

Error Measurement: the error measurement reported for Verbal, Performance, and the Full Scale are approximately equivalent to 3 IQ points, the correlations being significant at the .05 level.

Validity:

Face Validity: the face validity of the WISC-R appears quite appropriate with regard to its purposes. The subdivision of the Scale into Verbal and Performance tests seems an adequate way of identifying two major factors in human abilities. Indeed, the WISC is probably the most thoroughly researched and time-honoured standardized test in the world.

Content Validity: it is difficult to evaluate because the author provided a definition of the concept of intelligence which is so general that it cannot be directly related to the content of the test. Wechsler broadly defines adult intelligence as "the overall capacity of an individual to understand or cope with the world around him." He also mentions the presence of "non-intellective factors" that have been taken into account in his selection of a wide variety of types of items. However, no operational definitions of these factors are included, leading one to conclude that the WISC-R, like all other IQ tests, reflects the rather arbitrary conception of intelligence held by its author.

Concurrent validity: the Manual reports fairly high correlations between the WISC-R and the Stanford Binet IQ's within homogeneous age groups: the mean correlation with Full Scale IQ is .73, the Verbal Scale correlation rate is .71, while the Performance Scale correlation rate is .60. Correlations are also reported between the WISC-R, WAIS, and WPPSI. As can be expected in view of their authorship, the correlation between the WISC-R and WAIS Full Scale is .95, the correlation between the two Verbal IQ's is .96, and the correlation between the two Performance IQ's is .83. Similar high correlations are found between WISC-R and WPPSI: Full Scale is .82, Verbal IQ's .80, and Performance IQ's .80.

Predictive Validity: little evidence of the value of the WISC as a predictor of subsequent behaviour has been reported. As with the WISC, the WISC-R gives no criterion predictive validity data, despite the validity research available.

Construct Validity: closely related to content validity, the analysis of the subtests leads to the assumption of a general intelligence factor basic to the development of the items. Concerning "non-intelligence factors", there is no explicit statement of what these are or how they should be dealt with in interpreting subtest scores. Factor analyses on the WISC (1948) identified four factors underlying the test, but the most important finding was that Verbal and Performance Scales are significantly loaded with the verbal intelligence factors. Subsequent studies have reported similar findings in the identification of factors having substantial correlation with verbal components. Studies on construct validity of the WISC raise the question again on the lack of well-founded rationale underlying the test's development.

General comments: The WISC-R represents an improvement of the WISC, significantly minimizing the major deficiencies of the WISC while retaining most of its virtues.

The purpose of the revision, stated in the WISC-R Manual (1974), was to modify or delete ambiguous, obsolete or unfair elements, yet retaining as much of the original version as possible. Major changes in content, inclusion of non-white norms, a more active examiner's role, and improved scoring criteria are particularly noteworthy. However, despite some significant changes in administration and content, the WISC-R retains the fundamental strengths, weaknesses and utility of the WISC. The WISC-R still provides Verbal, Performance, and Full Scale based on the 12 subtests of the WISC, although the content or administration of most of the subtests has been modified. Item content has been modified, substituted or deleted on many subtests. Content has been altered to update obsolete material, eliminate uncommon or culturally unfair items, make the items more child-oriented, include pictures of non-white children and eliminate items that have proven too difficult even for adolescents. Another important change has been the modification of the sequence in which the subtests are administered, with Verbal and Performance tests given in alternating order, providing variety and reducing the risk of boredom. The direction and administrative procedures are a distinct improvement over the WISC. The new Manual and record form provide much more explicit and comprehensive instructions for starting points, discontinuance points, scoring criteria, and examples. The role of the examiner is much more active and demanding: the WISC-R requires the examiner to probe, demonstrate, correct, and otherwise actively evaluate the true extent of the child's knowledge and ability. Although this is a much needed improvement, it does place a greater burden of subjective judgement and technical competence on the examiner.

Reliability coefficients are as high or higher than those of the WISC, and are now given for all age groups, rather than for only the three in the original version.

The standardization procedure of the WISC-R includes a proportional representation of non-white children in the normative sample, but separate norms of the performance of white and non-white groups are not given. The Manual states that efforts were taken to eliminate culturally biased items, but there is not enough detail of the procedures which test developers followed to accomplish this. Although the comments of WISC users were solicited in revising items, we are not told how this information was employed. More importantly, although tryout samples were used for item revision of preliminary forms of the WISC-R, no information on the racial composition of tryout groups was given. It is important to note as well that in the standardization sample, mentally retarded, learning disabled and emotionally disturbed children, and groups with other handicaps were excluded. Yet this test is used primarily for the diagnosis of such handicaps.

The major weakness in the WISC-R Manual is the poor discussion of validity. Because of the lack of adequate content, construct and predictive validity, recent studies have questioned the usefulness of the WISC-R in making classifications of learning disabilities when used with ethnic minorities or socio-culturally different groups.

Earlier criticism of the WISC resulted in the reordering of items in terms of difficulty and objectivity of scoring in the revised edition, but the Manual does not present data on difficulty or on inter-rater agreement. It has also been a matter of concern that, even though the Manual provides the intercorrelations of all the tests by age group, no factor analyses are presented in support of the basic two-factor constructs which underlie the test.

Some reviewers have cautioned against the rise of the Verbal IQ, the Performance IQ or the subtest scores for diagnostic purposes. The Manual indicates that the WISC-R lends itself to psychological diagnosis and assessment of mental retardation, organic brain diseases and specific learning disabilities. However, the manual does not refer to the extensive literature discussion these use, and which generally report the test as being inadequate.

Concerning the assessment of culturally different children, the present norms appear to represent the national population well, according to the variables selected to define that population. But the inclusion of these groups in the general norms does not establish the validity of test scores for classifying members of various ethnic or socio-cultural populations, no matter how well the sampling has been carried out. Attention was given in the 1974 revision to eliminating items reported to be unfair or weighted toward the white middle-class culture. However, the WISC-R remains a test of the verbal and cultural components of the white mainstream society, and its use with ethnic minorities remains contentious. Studies on the WISC-R report high correlations between socio-economic status and test results, the scores obtained by high socio-economic status children being significantly higher. It appears that there is a need to examine present practices in test use with populations different from those in the standardization sample.

While the WISC-R is probably the most respectable standardized test instrument currently in use, it contains major conceptual and practical flaws. It is a good, reliable measure of those aspects of human performance which it taps, for those children who are neither culturally nor intellectually different from the population mean. The skills tapped have not been proven to be the main factors which predict school success, let alone the ability "to cope with and understand the world", which Wechsler sought to measure. More research is required before this test is adopted as a criterion test for intellectual performance, and especially for diagnosis of exceptional children.

Validity of the WISC-R as an Instrument for the Diagnosis of Learning Disability:

The Dubious Rationale for Using the WISC-R as a Diagnostic Tool: some attention has been given to interpreting gross differences between the Verbal and Performance IQ profiles as indicating the presence of a learning disability. Kirk and Bateman (1962) define specific learning disabilities as an educationally significant discrepancy between their estimated intellectual potential and actual level of performance. In this view, a learning disability is seen as a retardation or delayed development in one or more of the processes of speech, language, reading, writing, or arithmetic, which results from factors other than emotional disturbance, mental retardation, sensory deprivation, cultural, or instructional factors. Although these children meet a prerequisite criterion of "normal" intelligence, they show a variety of symptoms and conditions that are attributed to minimal brain dysfunction. The learning disability often is known to be and sometimes assumed to be an expression of a cerebral disorder. These may include dyslexia, dysphasia, dysgraphia, organically driven behaviour, distractibility, memory deficit and the imperception of visual, aural, or tactual stimuli.

McCarthy and McCarthy (1969) maintained that "identification of children with learning disabilities reduces to eliminating children whose behaviour can be explained on a basis other than cerebral dysfunction. For the children surviving this elimination some sort of cerebral dysfunction is presumed to explain the learning deficit."

Analysis of Discrepancies in Subscales, Verbal and Performance: the assumption underlying Verbal and Performance discrepancies is that both scales are tapping different brain functions. Verbal subtests are assumed to be dependent on language-related abilities (left hemisphere functioning), whereas success on the Performance subtests are assumed to be largely dependent on visual-spatial and visual-motor abilities (right hemisphere functioning). Thus, significantly low scores in either Verbal or Performance Scales, relative to performance on the other, is assumed to indicate deficits in one or another hemisphere of the brain. In children with learning problems, these deficits can also be manifested in difficulties in reading, writing, mathematics or other school subjects. Thus, a child who presents a discrepancy between Verbal and Performance scores is suspected of being learning disabled, and therefore of presenting a lateralized dysfunction of the cerebral hemisphere that will differentially affect his learning.

The WISC-R Manual provides statistical criteria for evaluating the significance of Verbal-Performance IQ differences, and such empirical guidelines are crucial for the interpretation of test results by clinicians. As the Manual states, to find an abnormal performance, "a difference of 9 points is required at the 15% level and a difference of 12 points is needed at the 5% level." These criteria are not substantiated and call for further investigation. Additional data pertaining to verbal-performance differences are needed for an adequate interpretation of these discrepancies.

A research study by Kaufman (1976) examined discrepancies between Verbal and Performance scores on the WISC-R, using the large and heterogeneous standardization sample as the data source for the study. Kaufman found that the mean absolute discrepancy score for a normal child is about 10 IQ points. Half of the children in the standardization sample had discrepancies of 9 or more points, about one-third had discrepancies of 12 and more, and about one-quarter had discrepancies of 15 or more IQ points. Furthermore, superior scores on the Verbal, as compared to the Performance, Scale occurred about as frequently as superior scores on the Performance, as compared to the Verbal, Scale.

Chronological age and Verbal-Performance discrepancy scores were similar for all groups tested, and neither sex nor race contributed to scoring patterns. However, parental occupation and the levels were statistically significant predictors of intelligence scale profile discrepancies. Children of professionals tended to score higher in the Verbal Scale, as compared to children of semi-skilled and unskilled workers, who tended to score higher in Performance. The mean discrepancy score was 10.7 for children of professionals and 9.2 for children of unskilled workers. The variation of the means for children in different IQ categories ranged from 10.9 for IQ's of

110-119, to 8 for IQ's below 80.

These findings, in particular, are of importance in the interpretation of test results when used with children presenting learning problems. Examiners must be cautious in interpreting various discrepancies in the child's patterns of test scores as abnormal, resulting in the diagnosis of a neuro-perceptual deficit or learning disability. A minimum of 12 points' discrepancy should be used as an index of relative weaknesses in cognitive functioning. More importantly, a low Full Scale IQ may not truly represent the child's intellectual functioning, in view of the discrepant patterns which might contribute to it. The results further indicate the need for a differential approach to the study of an IQ profile, depending on the child's background. A study of the frequency of children's Verbal-Performance discrepancies would be of practical importance in establishing a "normal" rate of variation among individuals.

Analysis of Subtest Profiles: Kaufman (1976) also investigated two indices of test scatter: the scaled-score range (a child's highest minus his lowest scaled score) and the subscale deviations (a spread of three or more points between a subtest score, the Verbal or Performance score and the overall mean scoring). The source of data was once again provided by the WISC-R standardization sample (n = 2,200). The results indicate that the scaled-score range for the Full Scale was 7.0 scale points, and the SD was 2.1. In other words, two-thirds of the sample had a scaled score range of between 5 and 9 points. It is evident that a typical child normally presents a fairly high scatter in his subtest profile. The high degree of scaled-score range was found throughout the sample, regardless of sex, race, age, socio-economic status, or intelligence level.

On the second measure of discrepancy between the Verbal and Performance subscale means, it was found that the average child usually had one test score that deviated significantly from his own Verbal or Performance mean, and two test scores deviating significantly from his overall test mean.

The implication of these results is that discrepancies within subtest performance are commonly seen in WISC protocols of normal children, and that these discrepancies are significantly large in about one out of three children. Obviously, diagnosis of specific learning problems, based on subtest or subscale discrepancies, can be highly suspect. Test scatters should not be considered alone as indicative of a learning disability, but should at most be interpreted as indicating a relative weakness in the abilities tapped by a given test. Consequently, if a child exhibited an unusual amount of test score scatter, there may be diagnostic and remedial implications. But it is important for the assessor to keep in mind that the normal child can be expected to demonstrate a very uneven subtest profile.

In the same vein, reseachers have so far failed to find significant differences in Verbal-Performance Scales between learning disabled and normal children. For example, Black (1973) reports that the achievement test performance and intellectual skill of reading disabled children with demonstrable neurological dysfunction appeared similar to those of reading disabled children without neurological dysfunctions. Bannatyne (1974) and Keogh et. al. (1974) attempted to combine or group subtest patterns as discriminators of perceptual disability and to correlate them with specific academic dysfunctions. Their purpose was to identify those problem-solving functions that influence school performance. However, Hullsman (1970), reviewing 20 previous studies, indicates that there is no evidence of a characteristic pattern of performance in learning disabled readers and suggests that research should be oriented to the analysis of the possible significance of low subtest scores instead of searching for predictive subtest patterns. Hullsman suggests that this approach would be more meaningful in terms of instructional programming for individual children.

Miller et al. (1978) aimed at evaluating the utility of the WISC patterns proposed by Bannatyne, Keogh and the Verbal-Performance model in the diagnosis of perceptual dysfunction. They assessed 121 learning disabled children; 73 of them had a primary visual perceptual deficit and 48 had a primary auditory perceptual deficit. It was found that: (a) "with each pattern more than half of the cases diagnosed as having an auditory perceptual learning disability were misclassified, that is, more than half of the auditory learning disabled subjects showed a WISC profile that could not be discriminated from visual learning disabled subjects"; (b) subtest profiles were not found to be associated with type of learning disability, that is, children who presented a learning process dysfunction in the visual or auditory mode did not necessarily exhibit an associated learning disorder. Similarly, children who showed a cognitive-level deficit, as measured by a WISC subtest score, did not necessarily show evidence of a previous perceptual deficit; (c) as the author states, "a WISC discrepancy between Verbal and Performance Scores was not found to be greatly indicative of a learning disability. Neither did results indicate that Verbal subtests were more affected by an auditory perceptual dysfunction or Performance subtests by a visual perceptual dysfunction." It is essential to explore the relevance of perceptual abilities as measured through psychometric tests to academic learning and school achievement. At the present, despite contradictory findings, it appears that there is little support for the assertion that particular deficits in perceptual abilities are responsible for poor academic achievement in specific school subjects.

In conclusion, the WISC-R is a limited tool for assessing cognitive functions in normal children as well as in children experiencing learning problems. The WISC-R should be administered as part of a battery of tests, but never be used as the only basis for diagnosis and programming.

Title of Test: Denver Developmental Screening Test (DDST)
Revised 1975 Edition

Author: William Frankenburg and Josiah Dodds

Publisher: Ladoca Project and Publishing FoundationInc.

Purpose: Diagnosis of Developmental Deviations

Type of Test: Individual Administration

Population for Which Designed: Children from Birth to Six Years of Age

Time Required To Administer: 15 to 20 minutes

Nature of Content: The DDST consists of a kit of materials (skein of red wool, a box of raisins, a rattle, a bottle, a bell, a tennis ball and eight cubes), a score sheet, and a reference manual. It evaluates four aspects of a child's functioning; gross motor , fine motor, adaptive language, and personal-social development. The DDST form is made up of 105 tasks or items written in the range of accomplishment of children in the age span from birth to six years of age. Items are grouped taking into account children's pace of development at different age levels. Age groupings cover a shorter time span at the younger ages, becoming gradually more distant as the test covers older ages. The test generates two types of scores: a total mental age score and a second score obtained from each separate subtest or sector. The results in each sector are categorized as "normal", "abnormal", and "questionable".

Practical Evaluation: Test materials are attractive and age appropriate - the test manual provides extensive instructions on administration, scoring, and interpretation of results. The test can be easily administered by non-professional examiners. The author claims that "with a few hours training any adult can administer the test competently" (Frankenburg et al., 1971).

Technical Evaluation: Norms: The Manual provides information on Standard procedures - the DDST was standardized on 1.036 (543 male, 493 female) normal Denver children between the ages of two weeks and 6.4 years. The sample excludes children who were adopted, premature, twins, breech deliveries, and who had serious handicaps of vision, hearing, central nervous system, cleft palate, etc.. The sample represents the ethnic and occupational groups of the Denver population according to the 1960 census. Items of the scale were selected after a pre-test of 200 infants and pre-school children and organized according to the age at which most of the children are able to do the task.

Validity: Concurrent Validity: The 1975 Revised Edition of the Manual presents data on concurrent validity based on 236 children who were administered the DDST and Stanford Binet, the Revised Bayley Scale for Children and the Revised Yale Developmental Schedule (RYDS). The correlations ranged between .84 and .97. Results were analyzed in terms of the probability that the screening and criterion tests will call a subject normal vs abnormal. The author reported that grouping the "normal" and "questionable" DDST ratings and all quotients of 70 and above together, resulted in identification rates of .73 and .92 with 7.2% over-referrals and 2.95% under-referrals.

Werner (1972), commenting on the results, says that "the analyses show a marked underselection by the DDST of children who received abnormal scores on the Bayley and Cattell tests." Seventy-one percent of the children called abnormal on the Bayley in the first year, and 43 percent of the children called abnormal on the Cattell in the second year, were called normal on the DDST - Children over 30 months who were classified as abnormal on the Revised Yale Schedule and the Stanford-Binet were generally correctly identified on the DDST. There was a considerable overselection of normal children in the third year of life: 42 percent of normal children on the Revised Yale Scale were called abnormal on the DDST.

There are almost no studies on the predictive validity of the DDST showing the relationship of results obtained in the scale and child behaviour several years later. Frankenburg carried out a study in 1974 on predictive validity but the results are as yet unavailable.

Reliability: The Manual presents data on test-retest reliability and reliability among examiners. Test retest reliability is reported as high as 95.8% agreement over a one week interval for a total of 2 months to 5 1/2 years old. For each child the percent of items performed the same way one week later ranged from 90% to 100%. However, a one week span is hardly long enough to assess the long-term reliability of the test.

Reliability among examiners based on the examination of 12 children is reported in the Manual. The percentage of agreement between four examiners on items passed or failed ranged from 80 to 95, with the average percent agreement being 90. A more extensive study carried out by the author and cited by Werner (1972) on interrated reliability of 76 children, ranging in age from 4 to 77 months, and a test retest reliability of 186 children between the ages of 1.5 and 76 months, reported co-efficients ranging between .66 and .93. However, in pointing out some of the weaknesses of the findings Werner stated that "reliabilities were generally poor in the first and second years, but reached a more acceptable level in the third and fourth year. Of some concern is the fact that of the 28 items with the highest tester-observer and test retest reliability, 17 could be passed by report of the mother, in contrast to only 5 of the 23 items with low tester-observer and retest reliabilities. The majority of the items on the DDST based on actual observations of the child's behaviour had only low to moderate reliability. Thus, the DDST seems to yield little more in the nature of reliable information than an interview with the mother, with the inherent distortion of development facts by recall."

General Comment: The Denver Developmental Screening test aims to be a short, practical sale, easy to administer and evaluate for children with serious developmental delays. However, the DDST presents weaknesses regarding norm groups, reliability and validity. The sample of children employed in standardizing the scale is not representative of the population to which these procedures are applied. Moriarty (1972), reviewing the test, states that "in our clinical use of the DDST, we had some reason to question the circumscribed geographical selection of the sample and the applicability of norms in screening children from the lower socio-economic groups, especially in the language area." Moreover, in view of the procedures followed to obtain the sample, the representativeness of the sample of Denver children appears somewhat doubtful.

Validity and reliability of the DDST are also questionable as mentioned above. The validity of the DDST relies on correlations with concurrently administered other tests or ratings by pediatricians, parents or teachers. Some of the results of validity studies are not encouraging. In one study (Frankenburg et al., 1971), it was found that the scale mislabelled normal children over the age of 30 months as abnormal; 42% of the children scoring in the normal range on the Revised Yale Developmental Schedule (RYDS) were called abnormal on the DDST.

Further studies also have shown a relatively poor concurrent validity of the DDST when checked against other IQ tests, as for example the Stanford-Binet.

The reliability of the DDST is based on test-retest and inter-examiner studies from small samples and heterogeneous age ranges that were tested at different intervals or on yet unpublished studies. A matter of concern is the sensitivity of the test to examiner's differences. A screening instrument is as good as the sensitivity of its user to the behaviour of young children. The interpretations of non-verbal clues from facial expressions, eye movements and gestures as well as the rapport established between adult and child and examiner's style varies according to amount of training, practice and examiners' background in child and developmental psychology. The author claims that one of the advantages of the scale is the fact that it can be administered by relatively untrained personnel with only a few hours training. However, later publications (Frankenburg et al., 1971) underline the importance of periodic retraining and repeated proficiency checks of examiners. Werner (1972) reports the findings by Furino and Connor in an unpublished paper in 1970 on comparisons made between the use of non-professional personnel referrals and referrals made by psychologists and pediatricians. It was found that non-professional personnel referred only about half of the children who presented developmental delays and only 1/5 of the children identified by pediatricians as having intellectual and developmental problems.

A further concern in the applicability of the DDST as a screening instrument is that its use has become over extended and its limitations disregarded by examiners. The author cautions against using the DDST as a diagnostic tool. He expresses in the manual (Frankenburg, 1975) ..."the DDST is a screening test, not an IQ test nor a definitive predictor of concurrent or future adaptive or intellectual ability." "The DDST alerts the professional to possible problems. It does not generate diagnostic labels such as learning disability, language disorder or emotional disturbance."

However, the DDST has become commonly used in the routine examination of young children, often by examiners who may not be aware of the author's cautions.

Given the questionable validity and reliability of the Scale, misdiagnosis and incorrect labelling could result. Furthermore, the predictive validity of the infant tests is highly questionable. Lack of correlation has been found between test results in infancy and later school ages. The evaluation of the sensorimotor abilities of the infant may not predict later intellectual abilities such as the comprehension and use of symbols and computer cognitive functions. Moreover, as Anastasi (1976) observed, infant scales "should be used principally to assess current developmental status rather than to predict subsequent ability levels. Development of abilities at these early ages is susceptible to so many intervening influences as to render long-term predictions of little value."

It should be kept in mind that the DDST is not as reliable, valid, and sensitive an instrument as the author presents it. Its reliability is doubtful for children under the age of 30 months, since it tends to underselect developmental problems. On the other hand, the scale tends to overselect normal children at the age of three years and on. In those age ranges where the DDST seem to be a satisfactory screening device, its concurrent validity is lower compared to other existing screening tools.

In conclusion, the DDST should be limited to use as a screening instrument, in the hands of skilled examiners, to assess current development only. The results obtained can help in alerting professionals of developmental delays, neurological defects, and environmental deficits that can be further investigated, but cannot predict long-term learning disabilities.

Title of Test: Coloured Progressive Matrices
Revised Order 1956

Author: J.C. Raven

Publisher: H.K. Lewis & Co. Ltd.

Purpose: "To assess the degree to which a person can think clearly on the level to which his intellectual functions have deteriorated" (Raven, 1965 Guide).

Type of Test: Group or individual administration

Population for Which Designed: Children from 5 to 11 years old
Handicapped subjects and senescents

Time Required to Administer: Maximum performance. No time limit.

Nature of Content: The test consists of three sets (A, Ab, B), composed of 36 matrices or designs from which a part has been removed. The subject chooses the missing insert from six given alternatives. Set A and B are identical to Raven's Standard Progressive Matrices but brightly coloured. An extra set Ab was designed to disperse the scores of children from 3 to 10 years of age and the mentally defective or impaired.

Practical evaluation: The test is easy to handle, to administer, and to score; it can be used by untrained personnel. In the form of boards, the test can be administered to persons from other cultures. It can also be used with people suffering from partial paralysis, deafness or defective speech. The manual provides fairly detailed information on administration, scoring and interpretation of results for the different populations. The score of the CPM is obtained from the simple addition of the scores on the three sets. This score is expected to be transformed in percentiles. Norms are presented only in percentiles.

Face validity: The task of choosing the part that best completes a design cannot be considered as a sufficient measure of a person's capacity for intellectual ability.

For this reason, the validity of the CPM can be questioned on the basis of a superficial evaluation.

Norms: The norms are based on 608 Scottish children for the book form (age range 5-11, roughly 50 per 6 months age group) and on 291 Scottish children for the board form (age range 5-10, roughly 32 per 6 months age group). The sample included children suffering from mental disabilities. There is little information about the characteristics of the children included in the sample.

Reliability: The author presents reliability coefficients for the CPM in terms of internal consistency and test-retest. According to Raven, for children under 7 years of age, the book and board forms of the test have shown a low retest reliability, in the neighbourhood of .65 and the correlation of about .50 with both the Crichton Vocabulary Scale and with the Terman-Merrill Scale Form L. By the age of 9, the retest reliability of both board and book forms of the test has been found to increase to at least .80 and to correlate about .65 with both the Crichton Vocabulary and Terman-Merrill scales. Over the whole range of development for which the test is constructed however, sets A, Ab, B, printed in colour, have shown a retest reliability in the neighbourhood of .90. These last results have been obtained from a limited sample of 25 normal and 29 disturbed children, from 6½ to 12½ years of age, who were given the Matrices and the Vocabulary Scale three times at intervals of 3 months.

No standard error of measurement is mentioned.

Validity: Construct and content validity: There is no clear rationale underlying the test. Raven's definition of mental development as expressed in the manual is that "mental development in childhood is more like salmon leaps in the stream of life than the equally arranged rungs of a ladder." But the nature of this process, or in other words, Raven's operational definition of intelligence is not specified anywhere in the manual. However, the author states that from the experimental work leading to the construction of the Coloured Matrices and from the examination of problems showing a low correlation with the scale, five stages of qualitative developments in the order of intellectual activity could be distinguished. The examination of these stages of development makes clear that they rely heavily on perceptual processes. Moreover, throughout the manual, Raven will refer to the CPM either as an instrument to assess a person's capacity for "productive thinking" or a test of "perceptual reasoning". This ambiguity is the origin for the lack of a definition in terms of the processes that are being measured by the test. It appears, generalizing from the Standard Progressive Matrices, that items are an attempt to measure factor "g" through perceptual reasoning. Wiedl and Carlson (1976) factor analyzed the structure of the CPM using a sample of 180 first-, second- and third-grade children. Their results confirmed the heterogeneity of content of the CPM test. Three orthogonal factors were generated that were interpreted as:

Factor I Concrete and abstract reasoning.

Items	Raven's classification
Ab8, Ab9	Discrete pattern completion of related figures as a whole
B8, B9	Concrete reasoning by analogy
B10, B11, B12	Abstract reasoning by analogy

Factor II Continuous and discrete pattern completion

Items	Raven's classification
A2, A3, A5, A6, A7	Simple continuous pattern completion
Ab2, B1, B2	Discrete pattern completion
B3	Discrete pattern completion or related figures as a whole

Factor III Pattern completion through closure

Items	Raven's classification
A4	Simple continuous pattern completion
A9, A10	Pattern completion showing progressive change
Ab3	Discrete pattern completion
Ab5, Ab6, Ab7, Ab8, B3, B4, B5	Discrete pattern completion of related figures as a whole

Despite the fact that the author was concerned with perceptual reasoning in the elaboration of the test, he does not present any evidence to support his hypothesis on developmental stages of perceptual processes.

Concurrent and predictive validity: The Coloured Matrices were checked against the Crichton Vocabulary Scales and the Terman-Merrill. For children 9 years of age, the correlation of CPM and Terman-Merrill is .66± .06, and CPM and Crichton Vocabulary Scales is.65± .07. For children between 6½ and 12½ years of age, the correlation between CPM and Crichton Vocabulary Scales is approximately .88 for a sample of 25 normal children and .65 for a sample of 29 emotionally disturbed children.

Raven concludes that the correlations are thought sufficiently high for both the Matrices and the Vocabulary Scales to be regarded as providing reliable long-term predictions of mental development. Despite the author's assertion, predictive validity cannot be considered as adequate due to the small size of the normative group and the lack of descriptive information about their characteristics. Although much has been done by other investigators to prove the validity of the Standard Matrices, this was not the case for the CPM. However, a recent article by Bentley (1976) reveals "statistically significant intercorrelations among RCPM and school achievement as measured through the Metropolitan Achievement Test (Reading and Mathematics)." No significant correlation between CPM and socio-economic level was found.

General Comment: The Raven Coloured Progressive Matrices have been extensively used and recognized as a useful instrument. It is easy to administer, requires little verbal instruction, is relatively brief and attractive to children.

The norms at present available for this test are unsatisfactory, calculated from a small sample. The scores obtained are transformed in percentile points.

Due to the fact that there is no full description of the group on which the data is based and that there are no derived scores equivalent for each raw score, the interpretation of results is uncertain. The author considers raw scores ranging only from 12 to 19 for 5½-year-old children, and from 17 to 35 for 11-year-olds, with 2 to 4 points of interval between classes. As a consequence, a difference of 1 point may signify the classification of a 6-year-old child as either "intellectually average" (raw score 13) or "definitely below average" in intellectual capacity (raw score 12). The implications for diagnosis can be serious. Raven himself is sceptical of the predictive value of the tests of children below the age of 6. He remarks that "between 3 and 6 years of age, a child's interest and attention are usually too mobile and his output of intellectual activity too capricious for any mental test to be a highly reliable predictor of his subsequent intellectual development."

The most serious criticism concerning the CPM is the validity aspect. The test does not have a clear formulation on the theory behind its construction. The author himself refers to the test as a test of relational thinking or perceptual reasoning. CPM would measure "g" through various forms of perceptual reasoning. However, the relevance of subcategories of perceptual functioning to intellectual functioning in general remains to be documented.

Morton Bortner, reviewing the Progressive Matrices in the Sixth Mental Measurements Yearbook, argues that:

> The use of perceptual items, however finely the author breaks them down into subcategories seems to provide a measure of perceptual adequacy rather than of intellectual capacity. The fact that perceptual adequacy and intellectual capacity are highly related in the normal, hardly provides a basis for using one as the measure of the other, especially in the abnormal.

Coloured Matrices have been constructed by introducing a new set of 12 items (Set Ab) in order to produce a wider dispersion of scores in children from 5 to 11. However, the insertion of the 12 new items seems not to have increased the scatter to any greater extent. The range and standard deviation of total scores obtained with the 1938 edition remains similar for the subsequent versions of the test.

Selection of items is also a questionable issue. There is evidence that the inclusion of too many easy items has resulted in a markedly skewed distribution which results in low validity. Furthermore, an examination of the test data on item difficulty shows that the items that form the different sets are not presented in increasing order of difficulty. For example, items Ab8 to 12 show a greater level of difficulty than do items B1 to B7 that immediately follow. This fact would work against Raven's claim that "the order in which the problems are presented provides the standard training in the method of working." In fact, this training is not provided because practice with a sequence of problems of one type would not necessarily help the child to solve problems of another type. Reliability of CPM is also a matter of criticism. Reliability with young children is not high (test-retest .65 with children under 7) and although the author claims a test-retest reliability of .9 over the whole range of development for which the test is constructed, the sample on which this association is based is relatively small.

In conclusion, despite the limitations indicated above, the Raven CPM remains as an interesting and valuable instrument to explore the possibilities of non-verbal testing. Its usefulness could be considered as part of a battery for the assessment of cognitive abilities for exceptional children, ethnic groups or children from low socio-economic background, providing the development of norms for these populations. However, much research needs to be conducted on theoretical and practical grounds to be able to obtain valid and reliable results from the CPM.

Title of Test: Slosson Intelligence Test (SIT)

Author: Richard L. Slosson

Publisher: Western Psychological Services

Purpose: Short intelligence screening test

Type of Test: Individual administration

Time Required to Administer: 10-15 minutes on average

Population for Which Designed: Infant to adult

Nature of Content: The test kit contains a manual and scoring sheets. It also includes the Slosson Oral Reading Test. The questions are presented orally to the subject and responses are primarily verbal at the pre-school level and beyond, with the exception of some motor tasks. The Basal Age is established when 10 consecutive questions are correctly answered. The test is discontinued when 10 consecutive questions are answered incorrectly. The score is then reported in terms of Mental Age with different weightings at different age levels. The IQ is obtained by dividing the MA by the CA times 100. An IQ Finder is available at the back of the manual. This produces a ratio IQ, not a deviation IQ.

Norms: The manual reports that the population on which it was standardized came from both rural and urban areas in New York State. However, it may be questioned as to how representative the sampling techniques were. The size of the population is not indicated.

Reliability: The reliability coefficient of .97 is reported for test-retest reliability on 139 subjects from age 4 to 50 years.

Validity: Since the SIT is based upon the Stanford-Binet, the concurrent validity is very high with the Stanford-Binet, yielding correlations between .90 and .97. The manual also reports that the SIT tests many of the same abilities as the Wechsler tests. The stability of IQ scores under the age of 4 is suspect and should only be considered as tentative.

General Comments:

There has been criticism of the statistical information provided in the SIT manual. The standardization group is not adequately described. The validity studies are unsatisfactory because of the small number of subjects (10 to 24) and because only average differences between the SIT, Stanford-Binet, or Wechsler are discussed. The fact that many Binet items are used produces correlations that are extremely high. Other studies report a wide variation of correlations between the SIT and Stanford-Binet (.76 to .90) and between the SIT and Wechsler (.54 to .93). Hunt (1972) expressed concern about the heavy verbal format between the ages of 2 and 4. She cautions that the "test validity is dubious for children with delayed language development and non middle class children." There is no mention of predictive validity in the manual. Hunt warns that "extreme caution should be used in relying on SIT test scores in situations where important diagnostic decisions are required...the amount of information derived from the test is limited."

Generally, the SIT is a rough estimate of intelligence functioning and serves the purpose of a screening measure but further assessment is required. Great caution should be used at the pre-school level as many factors may produce an unreliable score.

Title of Test: Illinois Test of Psycholinguistic Abilities (ITPA)
1968 Revised Edition

Author: Samuel A. Kirk, James McCarthy and Winifred D. Kirk

Publisher: University of Illinois Press, 1968

Purpose: The test was conceived as a diagnostic tool. Its objectives are: (a) to identify specific abilities and disabilities in children; (b) to provide basis for the development of instructional programs for children, founded on a final profile of the child's individual performance.

Type of Test: Individual administration

Population for Which Designed: Children from 2 to 10 years of age

Time Required to Administer: 45 minutes

Nature of Content: According to the manual, the ITPA is based on a psycholinguistic model that attempts "to relate those functions whereby the intentions of one individual are transmitted (verbally or non verbally) and reciprocally, functions whereby the environment or the intentions of another individual are received and interpreted." It deals with the psychological functions of the individual which operate in communication activities.

Based on Osgood's model of language, the test retains the elements of decoding (reception), encoding (expression], and association (mediation]. The authors operationalized the concept of psycholinguistic abilities by generating 10 discrete tests and two supplementary tests that evaluate the variables defining the concept. The 12 subtests are:

1. Auditory Reception
2. Visual Reception
3. Visual Sequential Memory
4. Auditory Association
5. Auditory Sequential Memory
6. Visual Association
7. Visual Closure
8. Verbal Expression
9. Grammatic Closure
10. Manual Expression
11. Auditory Closure
12. Sound Blending

Practical Evaluation:

The ITPA materials are generally interesting to children and the manual provides very specific directions for administration, scoring and interpretation of results.

The manual emphasizes both the need for a standardized procedure during the administration that must be followed closely and the need for a highly qualified examiner to obtain reliable results. The examiner should have an "adequate background" and should administer a minimum of 10 practice tests before being considered a valid examiner. The ITPA is a very difficult test to administer due to the many variables the examiner has to take into account in obtaining his or her results. In addition to the examiner's qualifications, the manual lists essential skills for reliable testing. The examiner must be aware of many variables such as directions to the child, timing, demonstration, need for maintaining the child's attention, motivation and co-operation, and the presentation of items in a spontaneous manner, within the framework of a highly structured administration format. Yet the examiner "should not be permitted to alter the directions by changing a phrase or omitting parts of the instructions."

Face validity: In spite of its title it is doubtful whether the ITPA contains measures of linguistic abilities per se. Specific subtests may measure the intended process, level or channel that they purport to measure, but their overall significance as linguistic components is questionable.

Norms:

Although the test is designed to identify departures from "normal", the standardization population, as in many tests, included only linguistically normal children. Authors report norms obtained on 700 children between 2 to 10 years old, who demonstrate average intellectual functioning, average school achievement, average personal-social adjustment, and who came predominantly from English-speaking homes. The standardization sample was drawn from five midwestern towns of moderate size in the U.S. and from homes slightly above the national average in income and education. This sample was evidently not intended to be representative of a North American population in general.

Reliability:

The manual contains no information on statistical procedures in the standardization and statistical characteristics of the ITPA. This information is provided by the authors in a separate publication (Papaskevopoulos & Kirk, 1969).

Internal consistency: The figures presented in the supplementary articles show acceptable internal consistency for all subtests through all age groups considered in the sample, especially for the subtests in the middle of age range. Empirical evidence generally seems to suggest that both composite and subtest scores are relatively highly correlated with each other, a positive characteristic of the test.

Test-retest reliability: The authors selected a group of children aged 4 to 8 years for testing and retesting after a five-month period. Children were drawn from the standardization sample, and examined in three groups of 4, 6, or 8 years of age. Test-retest coefficients were adequate for the composite score (.83, .70 and .70 for the three age groups), but test-retest stability correlations for individual subtests for periods up to six months are reported to be relatively low (six of the nine coefficients being less than or equal to .50).

Validity:

Strangely, the ITPA has proved to correlate better with standardized intelligence test batteries than with linguistic or psycholinguistic measures. Concurrent validity studies were carried out by McCarthy and Olson (1963), the criterion measures being language tests and achievement tests. The correlation coefficients for concurrent validity ranged from .03 to .60, while predictive validity coefficients ranged from .19 to .53. Several explanations were given for the low validity coefficients: (a) criteria tests and ITPA are not measuring the same abilities; (b) the restricted nature of the tests and the homogeneity of the ITPA sample would tend to lower the validity coefficients; (c) the subtests may not be measuring what they are intended to measure.

More recent studies correlating ITPA scaled scores and Stanford-Binet show generally high correlations of around .80. Hirshoren's (1969) study on predictive validity of the ITPA concludes that the test is at least as good as the Stanford-Binet in predicting intellectual achievement in "normal" children, although its validity as a measure of language has yet to be established.

Construct and Content validity: Studies have been carried out on the interrelations of the ITPA subtests and report that the ITPA's model of cognitive function does not really correspond to the theoretical model of Osgood upon which it was built. Uhl and Nurss (1970) subjected the ITPA results to factor analysis in an attempt to discover the factor structure underlying the various subtests. Only partial evidence of a single ability was found. The study, which used a large number of reference tests, was able to identify only three or four factors: (a) "vocabulary", which for upper middle class children was significantly tapped by at least six of the nine tests, but which split into "expressive" and "receptive" verbal factors for lower class children; (b) an "immediate memory span" factor present for both upper middle class and lower class children; and (c) a factor of "auditory processing", which was found only in the data for lower class children. The authors concluded that, considering the variables used in this study, the ITPA is primarily a measure of vocabulary skills for upper socio-economic status children and a measure of vocabulary and immediate memory for lower socio-economic status children.

Content validity: An examination of the subtests will lead the reader to the conclusion that, rather than a measure of psycholinguistic abilities, the items measure a "vocabulary" factor having to do with the range of cultural experiences to which the child has been exposed. Most of the items involve quite specific knowledge that can be expected only of children who have had wide cultural experiences. In certain subtests, the test items provide less information than they are expected to provide. For example, the test on Verbal Expression affords little information on the expressive linguistic development of the child by failing to assess particular measures that have been found to be highly indicative of level of linguistic development (e.g., mean length of utterance).

Predictive validity: Newcomer and Hammill (1976) have extensively reviewed the educational validity of the ITPA, using studies which included populations of children labelled learning disabled or retarded readers. Their criterion for judging the test as having predictive validity was a generous one; correlations of .35 or greater between test and criterion were considered acceptable. A summary of their findings follows.

Longitudinal Studies: Four studies reported the usefulness of the ITPA subtests as predictors of academic performance after a lapse of time (Egland, et al., 1970; Hirshoren, 1969; Mueller, 1969; Westinghouse Learning Corporation, 1969). Reviewing these, Newcomer and Hammill report:

> The conclusions of these studies are in agreement in reporting that only Auditory Association is a practically useful predictor of reading and that Auditory Reception, Visual Reception, Verbal Expression, and Manual Expression are not useful predictors of any academic variables.

Concurrent Studies: Newcomer and Hammill review 24 studies which attempted to establish concurrent relationships between various achievement tests and the ITPA. By reviewing the literature, a total of 1,152 correlation coefficients were obtained, most of which addressed the relationship between Reading and the ITPA. Newcomer reports:

> Only three subtests, Auditory Association, Grammatic Closure, and Sound Blending, as well as the Composite Score, established median coefficients which reached or exceeded .35. The other subtests including all those measuring visual processing skills yielded coefficients which are either not statistically significant at the .05 level or are so low as to have little practical value. (p. 50).

The relationships were even lower when intelligence was partialled out.

Grammatic Closure was the only subtest which showed any relationship to Spelling, and this was reported in only one study (Zbinden, 1970). Washington and Teska (1970) reported the only significant coefficients between more than one subtest and Spelling.

Concurrent studies for relationships between ITPA subtests and Arithmetic provide similar results as those obtained for Reading. Only Grammatic Closure, Auditory Association, and the Composite Score showed any correlation, and the one study which controlled for mental age (Hammill et al., 1975) showed no correlation between academic Arithmetic and any of the subtests.

Newcomer and Hammill (1976) summarize their findings as follows:

> Most significant is the indication that 9 of the 12 ITPA subtests lack predictive validity for any aspect of academic achievement studies. At best, only Auditory Association, Grammatic Closure, Sound Blending, and the Composite Score evidenced any predictive validity; however, when the influence of intelligence is controlled for, the coefficients associated with Auditory Association, Sound Blending, and the Composite Score drop below the established level for practical significance.

Educational Adequacy of the ITPA for Diagnosing Learning Disabilities:

Many authors point out that the ITPA is by far the most widely used test for identifying learning disabilities. Hallahan and Cruickshank (1973) suggest that "it is noteworthy that the ITPA, a primary if not the primary test identified with the field of LD, was developed in its original conceptual state within the area of mental retardation" (pp. 107-108).

Johnson and Morisasky (1977) report:

> Probably the most familiar learning disability model is Kirk and Kirk's (1971) approach to the diagnosis and remediation of psycholinguistic learning disabilities. The model is popular for several reasons . . . the diagnostic instrument intimately connected with this approach to working with learning disabilities, the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy & Kirk, 1968) has been widely used in both its original experimental and current revised forms. (p. 103)

Coles (1978) also states, "The Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy & Kirk, 1968) purports to measure those input, processing, and output abilities that underlie language development; it is perhaps the most frequently used learning disabilities test."

Given the popularity of the measure, evaluation of its cognitive and educational usefulness as an assessment instrument becomes necessary. In relation to this, Newcomer and Hammill (1976) contend:

> The principal value of any assessment device is that its use in the schools depends upon the extent to which it relates to academic achievement. No matter how astutely designed, accurately standardized, or brilliantly conceptualized a test may be, it is relatively worthless to educators unless its results can be used to improve children's capacity to perform such basic school activities as reading, spelling and arithmetic. In other words, it must be demonstrated that a test has educational significance. (p. 884)

The fundamental concept of learning disability is predicated on a malfunctioning in learning and subsequent failure to achieve full academic potential. The prevalent use of the ITPA with children suspected of having learning disabilities "suggests that many educators and psychologists have accepted as valid the premise that the ITPA is an educationally significant test" (Newcomer and Hammill, 1976, p. 844).

Research has indicated two major flaws in the ITPA which virtually destroy its capacity to provide a useful diagnostic instrument for exceptional children. First, it does not appear to measure any specific language skills, demonstrating, as reported above, very poor concurrent and construct validity, even with normally functioning children. Second, even if it constituted a valid measure of psycholinguistic ability, which it does not, it was standardized on a sample of children which was unusual by test development criteria. The highly selected sample excluded any child of abnormal IQ, school achievement, personal-social adjustment, sensory and physical handicaps, and any child from an ethnic minority second-language background or low socio-economic group. The authors report that of 2,413 children considered for inclusion in the norm sample, 58% were rejected. Such a sample makes questionable the authors' claim that the test norms provide a broad basis for diagnosing children with learning disabilities. As Carroll (1979) remarks, "it would appear, from studies reported in the literature that by far the greatest use of the ITPA has been in the assessment of the psycholinguistic abilities of lower class children particularly from minority groups: Blacks, Chicanos, and even Papago Indians." The most common use of the test is evidently with groups specifically excluded from the norm sample.

There have been some attempts to assess the usefulness of the ITPA in discriminating learning disabled children from normal children. Bateman (1968) has provided some tentative group profiles which she suggests can serve as norms for mentally retarded children, low socio-economic, sensorially impaired and physically handicapped groups. Harth (1977) examined the validity of the Visual Closure subtest with 94 severely learning disabled children. Also concurrently administered were the Frostig (Developmental Test of Visual Perception) and the WISC-R. The Visual Closure subtest poorly discriminated the children, and more closely correlated with figure-ground perception than with other measures of linguistic or perceptual skills, leading Harth to conclude that at least this subtest has limited diagnostic and predictive value for learning disabled children.

Larsen (1976) examined the diagnostic usefulness of some commonly used tests of perceptual functioning, among them the ITPA (Auditory and Visual Sequential Memory and Sound Blending subtests), the Wepman Auditory Discrimination Test and the Bender Visual-Motor Gestalt Test. Thirty normal and 59 learning disabled children were tested to determine if there were differences in performance between the normal and learning disabled children. Results showed that only the Bender-Gestalt differentiated between the two groups.

Several researchers have questioned the usefulness of the remedial procedures which McCarthy has developed to supplement the psycholinguistic profiles produced by the test. Foremost among these are Newcomer, Larsen and Hammill (e.g., Newcomer and Hammill, 1976), who, in various issues of the journal Exceptional Children, have parried with the test authors, at times in forceful terms. Werner, Barritt and Simmel (1967), working with learning disabilities, expressed a need for evidence to show that subscale performance on the test was related to specific educational behaviour. Favrier (1963), working with functional disorders of articulation, complained that the remedial procedures that would be suggested by test performances would in most cases bear little resemblance to what is usually considered speech therapy.

Carroll reports in studies conducted to ascertain the usefulness of the test in prediction or diagnosis of reading disabilities. The results are divided, with Bateman (1963), and Kass and Macione (1966) showing favourable evidence, while Sears (1969) found "no significant difference between average and disabled readers in any of the ITPA skill areas." Furthermore, Smith (1969) found no relationship between auditory and visual types identified by the test and method of reading instruction used. Thus, even when used on populations closely resembling populations included by a variety of learning disability definitions, the test's diagnostic and discriminatory abilities are highly questionable.

In summary, at the risk of overstating the case, the ITPA appears to be of little use as a measure of general linguistic abilities, let alone as a tool for discriminating the learning disabled child. It is not at all clear what the test measures, in spite of its title, its authors' claims, and the programs for remedial intervention which accompany it. Of greater concern is that in the hands of inexperienced assessors or gullible clinicians, the test could lead to serious misdiagnosis of a child's abilities, and result in inappropriate placement and remediation.

Title of Test: McCarthy Scales of Children's Abilities (MSCA)

Author: Dorothea McCarthy

Publisher: The Psychological Corporation

Purpose: Diagnostic profile of abilities and general intellectual level.

Type of Test: Individual administration

Population for Which Designed: 2½-to 8½-year-olds

Time Required to Administer: 45 minutes to 1 hour, depending on age

Nature of Content: The MSCA consists of a kit of materials (Card Book, 1-inch cubes, puzzles, xylophone, tape, rubber ball, bean bag, target, tube, pieces of cardboard, set of blocks that are round and square), a manual, record forms and drawing booklets. The 18 tests evaluate six different aspects of the child's functioning; Verbal, Perceptual-Performance, Quantitative, General Cognitive, Memory, and Motor. The Verbal, Perceptual-Performance Quantitative Scales constitute the elements that form the General Cognitive Scale. For each of the six scales, the child's raw score is converted to a scaled score, called an Index, according to chronological age. The General Cognitive Index (GCI) has a mean of 100 and a standard deviation of 16. The scores for the remaining five scales have an arbitrarily chosen mean of 50 and a standard deviation of 10.

The subtests are well integrated to provide a variety of tasks and novel stimuli making it an interesting test for both the child and the tester.

Norms: The manual reports that the standardization of the MSCA involved a U.S. nationwide sample and was stratified on several major variables according to the U.S. Bureau of the Census. These variables included age, sex, color, geographic region, father's occupation, and urban and rural residence.

Ten age groups were defined with 100 children at each level, making the standardization population of 1,032. Scaled score distribution for the six scales were determined for each of the 10 half-year age groups from 2½ to 8½ years.

Reliability: Reliability coefficients for each of the scales based on either split half reliability or test-retest reliabilities, whichever being appropriate, yielded .93 for the General Cognitive Scale and the average for the other scales range from .79 to .88. The standard error of measurement for the General Cognitive Index is approximately 4 points.

Validity: The MSCA correlates with the Stanford-Binet .81 for a sample of 35 6-year-olds. The MSCA correlated with the WPPSI at the .71 level and McCarthy notes that this lower correlation could be explained by the restriction of range in the WPPSI at the age tested. The predictive validity was obtained on a sample of 31 children aged 6 to 7, with the Metropolitan Achievement Tests administered 4 months after the MSCA, and this yielded correlations ranging from .34 to .54 using the GCI. Construct validity appears to be satisfactory as established by factor analytic techniques.

General Comments: The McCarthy Scale of Children's Abilities appears to have good reliability and validity. However, Sattler (1974) indicates that the validity of the Motor Scale needs to be established. He also warns "significant scatter in 62% of the profiles of the standardized sample has been reported", and means that caution must be used in interpreting scatter on the MSCA, especially with exceptional children. Kaufman and Kaufman (1977), in reviewing recent studies, indicate that with a learning disabled population, the MSCA gives a lower General Cognitive Index by approximately 15 points than the WPPSI and the Stanford-Binet. They also comment that "A clinician will not be able to diagnose a child as learning disabled based solely on a low GCI since mental retardation, emotional disturbance, sensory loss, etc., may be alternate explanations of poor performance on the McCarthy."

Basically the MSCA offers some important data on a child when used in combination with other tests. There are no reports that offer strong support for the predictive validity of school achievement using the MSCA.

Title of Test: Bender-Gestalt Test (BGT)

Author: L. Bender.

Publisher: Psychological Corporation, 1938

Purpose: Maturational test in visual-motor gestalt functioning in children.

Type of Test: Individual administration.

Population for Which Designed: Children to adults.

Time Required to Administer: Unspecified, but usually 10 to 15 minutes.

Nature of Content: The Bender-Gestalt consists of nine cards with a simple design on each card. The subject is requested to copy each design with a pencil onto paper, while the sample remains before him. The Koppitz (1964) method of scoring is used most frequently with young children 5 to 10 years old and was prepared as a nonverbal developmental scale.

Norms: The Koppitz (1964) method of scoring was developed following the administration of the Bender-Gestalt Test to 1,104 kindergarten to fourth-grade children in midwestern and eastern United States schools.

Reliability: Koppitz in his book reports interscorer reliability correlations between .88 and .96. Retest reliability over a four-month period produced low coefficients ranging from .547 to .659. The older the children, the higher was the retest reliability. Within single-year groups, correlations from .48 to .79 were found between the BGT scores and Stanford-Binet or WISC IQ's. After the age of 10, the Bender-Gestalt no longer correlates significantly with either intelligence or age. Bender cautions in her 1946 instruction booklet that the results can be adversely affected by fatigue, thus lowering retest reliability. Kitay (1972) points out that variations in instructions such as those of Hutt would seem to make a difference in performance, thus lowering retest reliability.

Validity: Koppitz (1964) reports fairly high predictive validity in assessing school readiness and the first-grade level of educational achievement. Studies of first and second grade children showed significant relationships between BGT scores and performance in reading and arithmetic. Koppitz attributes this relationship to the role of visual-motor perception in the early stages of learning these subjects. He further suggests that among mentally retarded children, the Bender-Gestalt developmental score has fairly high validity as a measure of intellectual level.

These analyses are based on small samples and need further verification.

General Comments: The Bender-Gestalt Test using the Koppitz developmental method of scoring is used to assess visual-motor performance for children aged 5 to 10 and to predict IQ and school achievement. The reliability and validity present weaknesses.

Kitay (1972), in reviewing the Koppitz method of scoring, expresses concerns about construct validity. Coy (1974) reports that the Bender-Gestalt had no significal predictive validity for predicting reading and math achievement for regular grade three pupils. Coy also warns that "BGT results alone should not be used to establish whether or not a visual-motor deficiency exists." Zach and Kaufman (1972) report that "test-taking" experience is a critical factor in raising scores. They retested black kindergarten students three weeks after first administering the BGT, and found that initial scores indicating impaired visual-motor performance were increased to within the normal range. This casts doubt upon the validity of the visual-motor component. In a second study, they report that among 70 kindergarten children from the lower socio-economic classes, "it was possible for a child to discriminate forms well and still obtain a score on the Bender which indicates 'perceptual' difficulties, and vice versa." Many researchers have previously criticized the BGT because it provides a global measure, in which not only visual perceptual skills, but also motor and other skills, are tapped. Furthermore, it fails to differentiate between the various perceptual abilities. Bauman and St. John (1971) compared the Bender-Gestalt, Frostig Test of Visual Perception, and the Minnesota Percepto-Diagnostic Test as measures of visual perception. Their results indicated that significant correlations between the tests were found for children at some ages and not at others. They concluded that these tests measure different perceptual functions which are independent of each other. On the more positive side, Thweatt, Obrzut & Taylor (1979) have developed an objective scoring system that purports to measure an individual's directional orientation to perceptual tasks. This "Soft-Sign System attempts to identify those 'mild' perceptual factors that significantly correlate with reading prediction." Their sample included 100 children aged 6 to 9 and 11 to 12 years. Results indicated that the Soft-Sign scoring system may be valid as a predictor of reading achievement. Further investigation is required on other age groups and socio-economic levels, as well as exceptional children.

Generally, the Bender-Gestalt, using the Koppitz scoring system, lacks adequate validity and reliability and should be used with great caution and only as part of a battery of tests to assess or predict a range of possibilities.

Title of Test:	The Screening Test for Identifying Children with Specific Language Disability (Slingerland) Revised 1970 edition
Author:	Beth H. Slingerland
Publisher:	Educators Publishing Service Inc.
Purpose:	To identify children with specific perceptual-motor disorders (visual, auditory, and kinaesthetic) that may interfere with adequate development of early reading, writing and spelling. The eight subtests are designed to reveal problems in the processes of association and interaction of perception, discrimination, integration, memory, and performance. The test is based on the Orton and Gillingham techniques for teaching reading through multisensory channels.
Type of Test:	Individual administration.
Population for Which Designed:	Children in grades one to four.
Time Required to Administer:	One hour, in two or three sessions.
Nature of Content:	Eight subtests; visual copying, far point and near point, visual perception - memory, visual discrimination, visual perception - memory in association with kinaesthetic memory, auditory recall, auditory perception of beginning and ending sounds, auditory association.
Technical Evaluation:	The author presents the test as a clinical diagnostic instrument, to be interpreted in the context of each child's learning history and those factors which affect his performance (limited learning opportunity, poor health, etc.); therefore, no norms were reported, and consequently no reliability or validity data. The author suggests that local norms may be developed as needed. Mention is made (1970) of a study of 10,000 children purportedly underway.
General Comments:	Deno in Buros (1972) criticizes the test author for lack of documentation, and lack of data on validity and reliability. Deno notes that while the clinical use of the Slingerland has value, the lack of a normative base seriously jeopardizes the validity of the test. Instead of norms, Slingerland provides a descriptive profile for normal and learning disabled children at each grade level in half year intervals, but since this is not empirically based, it must be accepted with caution. J.M. Wepman (1972) feels that the test's title is misleading for "no measure of language is included," and that the Slingerland test is complex in its organization,

administration, and scoring and requires too much study before use by the examiner. While Form A, grades 1 and 2, are in Wepman's opinion worthwhile, Form C, grades 3 and 4, are too simple to provide a fine differential diagnosis. The subtests involve two copying tasks, two word matching, one writing from memory, two writing from auditory presentation and one calling for visual recognition and encircling a response. Wepman considers that these tests might provide an adequate screening test for perceptual integration, but he is sceptical that the emphasis of the test on the perceptual process within the three modalities, rather than on conceptual processes, can predict success in reading. Wepman is also critical of the omission of tests of expression, oral or written, and of interpretation of perceptual, vocabulary or language patterns. In his view, the tests relate to pre-linguistic rather than linguistic behavioural tasks.

E. Deno (1972) sees the Slingerland test as comparable to the ITPA or the Peabody Picture Vocabulary Test and other tests of the assessment of language functioning, but the usefulness of these tests are in his view not clear. More empirical evidence of the effectiveness of the test is needed.

Donah's (1977) study aimed to provide validity for the use of the Slingerland test in identifying children with learning disabilities. Donah studied the test's construct validity with a rural, northeastern Ohio school population of grade one children from low income and welfare families. Twenty-one of the original twenty-six children were part of the final validation sample. Twenty children were suspected of having learning disabilities on the rank ordering of the Slingerland test. The canonical correlation indicated a significant relationship between the Slingerland subtest profile and performances on individually administered criterion test batteries of perceptual skills and classroom achievement.

Although this study involved only a small number of abnormal children, the results demonstrated that the Slingerland profile might be useful in identifying children with learning disabilities, as defined by their performance on specified classroom performance variables. Donah states that poor performers on the Slingerland were children who: (a) displayed perceptual-motor dysfunctions according to classroom teachers; (b) displayed more academic difficulties as indicated by performances on standardized reading tests and by grades in reading, math, and academic placement at the year-end, and on the Stanford Achievement Test. She concluded that the Slingerland has concurrent predictive and construct validity and that it can be used "with some confidence to select first-graders who evidence a high probability of learning disabilities."

Title of Test:	Developmental Test of Visual Perception (DTVP)
Author:	Marianne Frostig
Publisher:	Consulting Psychologists Press
Purpose:	Assessment of visual perception.
Type of Test:	Individual or group administration.
Population for Which Designed:	3- to 9-year-olds
Time Required to Administer:	30 to 45 minutes
Nature of Content:	The test materials consist of an expendable booklet, the back cover of which serves as a scoring sheet, 11 demonstration cards, and three transparent scoring tissues. The test is scored according to clearly defined criterion. Raw scores are converted to Scale scores with the use of tables. PA/CA X 10 = Scale Score. Perceptual Quotients (P.Q.) are also available for children age 4 to 8 and the P.Q. can be translated into Percentile Ranks using a table. The test is divided into five subtests of: Eye-Motor Co-ordination, Figure-Ground, Constancy of Shape, Position in Space and Spatial Relationship. The child receives a score for each test separately, then a total test score, reported as a Perceptual Quotient.
Norms:	The 1963 edition of DTVP was standardized on 2,116 children and the revised edition (1966) on 2,100 children. The samples were not drawn from a representative cross section of the population, but disproportionately represent middle-class children.
Reliability:	The manual reports test-retest reliability of the perceptual quotients to be +.80, for a small group of first and second-grade children. Test-retest reliability for the entire sample was .70. Subtest scale score test-retest correlations ranged from .29 to .74. Split-half reliability correlation coefficients for the total test were found to decrease slightly with increasing age.

Validity: Significant correlations were found between the performance of low scores on the DTVP and teacher ratings of classroom adjustment. Low correlations were found between the DTVP and the Goodenough Harris Draw-a-Man Test for children in Kindergarten to grade 2.

General Comments: A factor analytic study (Sabitino et al., 1974) indicates that the DTVP contains two principal components. One is visual-perceptual discrimination (position in space, spatial relationships and constancy of shapes) and the second component is manual motor co-ordination (eye-motor co-ordination and figure-ground relationships). Smith et al. (1972) suggest factor analysis of the test yields only a single general factor of perceptual organization. This is weakly related to IQ and unrelated to reading ability.

The standardization sample on which the test was validated has been criticized by many authors for poorly representing minority children and lower socio-economic groups. In addition, several authors (Allen et al., 1968) have noted that the test must be interpreted with great caution for children whose handicaps may include retardation.

The literature indicates that the DTVP predicts educational achievement best for younger children, prior to grade 3. Although the DTVP may predict reading problems in the early grades it does not predict specific kinds of reading problems or long-term disability. Mann (1972) in his review indicates that the DTVP "does discriminate between normal and neurologically impaired population", but he questions the power of the DTVP in predicting learning potential.

In summary, the DTVP seems to measure, at best, some aspects of perceptual motor organization in young children but with weak power for predicting school achievement. As Mann cautions, low scores on the DTVP would not be a valid indication of the need for perceptual training.

Title of Test:	Developmental Test of Visual-Motor Integration (VMI)
Author:	Keith E. Beery and Norman A. Buktenica
Publisher:	Gollett Publishing Co.
Purpose:	Assessment of visual-motor behaviors
Type of Test:	Individual or group administration
Population for Which Designed:	Children from ages 2 to 15 years
Time Required to Administer:	10 to 20 minutes
Nature of Content:	There are two forms of the test booklet. The short form for ages 2 to 8 is identical to the long form for ages 2 to 15, except that the last nine items are omitted. In the long form there are 24 geometric designs to be copied. The geometric forms are arranged in increasing order of difficulty. The manual gives well-defined, explicit criterion for the pass-fail scoring on each item. The Visual-Motor Integration (VMI) Age Equivalent is obtained by counting the number of correct designs up to three consecutive failures and consulting the VMI Age Equivalents table.
Norms:	The manual gives no information on a standardized population.
Reliability:	The manual does not provide any information on reliability.
Validity:	It is reported in the manual that the correlation between VMI scores and chronological age is .89 for the 2 to 15 age range, although the source of this data is not reported. The author indicates that VMI correlations are higher with mental age than with chronological age, and that VMI correlates with reading achievement in the first grade at a higher level than the correlations between IQ and reading achievement.

General Comments:

Leton (1974) in his review is critical of the omission from the test of graphomotor skills needed in printing and writing. He states that since the authors did not obtain drawing responses from a criterion group of children with graphomotor problems, the diagnostic validity of the VMI to identify learning disabled has not been demonstrated. Rice (1974) expresses concern about the validity of the test and calls for more research in this area. Basically, there is little research data available to support the use of the Developmental Test of Visual-Motor Integration in a testing battery. In view of the lack of normative information, this test cannot be viewed as an objective or valid assessment instrument.

Title of Test: Auditory Discrimination Test (1958)

Author: Joseph M. Wepman

Publisher: J. Wepman, Chicago

Purpose: The test was conceived by the author as "an easy to administer method of determining a child's ability to recognize the fine differences that exist between the phonemes used in English speech."

Type of Test: Individual administration

Population for Which Designed: 5- to 8-year-old children and older children who seem delayed in developing speech accuracy on an auditory basis

Time Required to Administer: 5 minutes

Nature of Content: The test consists of Form I and Form II, two equated forms permitting test-retest evaluation in short periods of time as well as providing a check for children who may have inappropriate scores on the first administration. According to the manual, each form consists of 30 pairs of words differing in a single phoneme in each pair and 10 word-pairs which do not differ as false choices. Comparisons are made between 13 initial consonants, 13 final consonants, 4 medial vowels and 10 false choices. Score is determined by the number of errors the child makes in calling words which are different identical.

Practical Evaluation: This is a short and easily administered test that does not require visual, speech, or reading ability to arrive at its results. The test provides a preliminary manual and scoring sheets. Administration does not require manipulation of materials. The examiner reads the word pairs and the child indicates whether the words in each pair are the same or different. The manual provides adequate instructions for administration and scoring.

Norms: The manual does not provide information on the normative sample and procedures followed to collect data. The author briefly mentions the testing of 533 unselected first, second, and third-grade children in both urban and non-urban communities in order to obtain criteria to judge the development of auditory discrimination. Based on the assumption that auditory discrimination is developmental, the author obtained a profile of errors at different ages which are used as norms for scoring purposes.

Reliability: A test-retest reliability is reported as .91 on 109 subjects.

Calculation of phoneme difficulty on the two forms of the test resulted in a Pearson rank order correlation of .67 for 214 cases.

Validity: Construct Validity: The manual does not provide a rationale for test development. In a later publication, Wepman (1960) states that "essential to the development of auditory discrimination is the ability to retain individual sounds in mind to serve as models for later speech and as part of the phonic act necessary for reading. Both discrimination and retention must reach a satisfactory level of development before the child can use them for accuracy in speaking or for word attack in reading." However, he does not clarify the relationship between auditory discrimination as measured in the test and reading achievement.

Concurrent Validity: Correlation coefficients of r = .32 are reported between the results of the auditory discrimination test. Results indicate a relative lack of relation between the development of auditory discrimination and intelligence as measured by most intelligence tests.

General Comment: The Auditory Discrimination Test is a quick, inexpensive and easy to administer and score device. It is widely used in schools to select children who are experiencing difficulties in learning to read, on the assumption that particular auditory skills are prerequisites for mastering reading. According to the author, the test "has been found useful in selecting children, especially those in the early elementary school years, who are slower than their peers in developing auditory discrimination." Auditory discrimination ability has been found to be highly related to the development of speech accuracy; however a poor correlation is shown between poor discrimination scores and poor reading scores (Wepman, 1960). Research in the area supports this finding, suggesting that auditory skills are not sufficiently related to reading to be particularly useful for school practice and do not seem to provide information relevant to academic learning. Larsen et al. (1976) carried out a research assessing 89 children from 8 to 10 years of age. Fifty of them were considered as learning disabled. Tests of intelligence and perception were administered. Among the latter, the Auditory and Visual Sequential Memory and Sound Blending Subtests of the ITPA, the Wepman Auditory Discrimination Test, and the Bender Visual-Motor Gestalt. The results showed that the difference in performance between the LD and normal children was not statistically significant, which means that a child who is performing adequately at school cannot be differentiated from the one who performs poorly on the basis of perceptual tests. Similarly, Hammill et al. (1974) reviewing studies that correlate measures of reading to measures of auditory discrimination, memory, blending, and auditory-visual integration conclude that it appears there is little correlation between auditory skills and reading as measured through the standardized clinical instruments frequently used as ITPA, Wepman's Auditory Discrimination Test, Belmont's Auditory-Visual Integration, etc. Therefore these instruments can be considered of little predictive value of reading achievement.

There seems to be a lack of definition or agreement as to what constitutes the basic "perceptual" or "psychological" processes involved in the process of learning. The concepts hypothesized to be causing the child's failure at school need to be clarified and defined to be useful tools in the diagnosis of learning problems. Concepts, for example, auditory discrimination, are defined and measured in different ways depending on the examiner's subjective notion of what the definition of the skill is and the manner of assessment used to measure that specific ability. Wepman's Auditory Discrimination Test can be criticized on the grounds of its construct validity. As Hammill et al. (1974) conclude, the existing studies correlating reading and auditory skills "fail to prove that particular auditory skills as measured, are essential to the reading process, and that many children actually fail to read proficiently because of auditory perceptual deficits. Apparently a large percentage of children who perform adequately on tests of auditory perception experience difficulty in learning to read; and an equally sizable percentage who do poorly on these same tests have no problem in reading."

The validity of the Wepman's Auditory Discrimination Test can also be questioned because of the lack of information on standardization procedures and normative sample. The author mentions small samples of children tested in Illinois and Florida, but there is little information concerning age, sex, IQ, or socio-economic status. This fact makes the test a questionable instrument to assess children from low socio-economic status or coming from different cultures. The data provided by the author on validity and reliability of the test is minimal and undoubtedly its construct and concurrent validity are poor.

In conclusion, the Wepman's Auditory Discrimination Test is a quick and easy device to administer to children from 5 to 8 years of age, who speak standard English, to assess current development of speech. The deficiencies of the test in terms of normative data, reliability, and most of all, on validity data, makes this instrument of doubtful value in assessing children at school who are experiencing difficulties in reading. The use of this instrument in circumstances where the concept of "perceptual" deficits as they relate to school achievement is controversial would only contribute to the labelling and misplacement of children.

Conclusion

While each of the preceding tests may have some type of validity in its own right, none of these tests were created for the purpose of identifying children with learning problems. Indeed it has not been established that the relationship between visual perceptual motor problems and poor academic achievement is a causal one. Therefore the use of these visual motor tests as predictive measures of learning problems with the pre-school population would seem highly questionable. This becomes evident when examining the literature specifically addressing the question of predictive validity concerning school achievement. At best, test results in kindergarten can predict only school achievement in grade 1 but seldom beyond, and this is based on minimally acceptable correlations and small samples. In conclusion, the use of the tests reviewed here for the early identification of learning problems cannot be endorsed.

Learning Disabilities Assessed Through Standardized Achievement Tests

In assessing children for learning disabilities, frequently low scholastic performance in one area of achievement is contrasted with higher performance in other areas. Identification of a learning disability has also been based upon a discrepancy between potential and actual functioning levels. In both cases, standardized achievement tests have been used as diagnostic instruments. In the case of the first criterion, disparate scores on subtests are used as indicators. In the case of comparing potential and actual functioning, a measure of intelligence purports to indicate the child's potential, while the overall achievement test results indicate his actual level.

How validly can tests of achievement, usually standardized on normal populations, be used as diagnostic tests for learning disability? Several studies have addressed this question, some examining the predictive power of a test in discriminating a learning disabled from a normal sample, others comparing the validity of one or more tests in correlating with each other and with criteria such as grade level and teacher ratings.

Several well-used achievement tests are reviewed in this section. In the first part, each test is considered alone, and in the second part, studies are reviewed, which explore the intertest and predictive validity of groups of tests.

Title of Test: Wide Range Achievement Test (WRAT)
Revised 1976 edition

Author: J.F. Jastak and S.R. Jastak

Publisher: Guidance Associates of Delaware

Purpose: Assessment of basic school subjects: reading, word recognition and pronunciation, written spelling and arithmetic computation.

Type of Test: Primarily individual administration with provision for group administration of some parts under specific conditions.

Population for Which Designed: 5-year-old children to adults

Time Required to Administer: 43 minutes

Nature of Content: The test consists of 3 subtests: reading, spelling and arithmetic, each one of them ranging from kindergarten to college. Each subtest is divided into two levels designed to be used with different age levels: Level I, for children 5 to 11, and Level II, for ages 12 and over.

Scoring is obtained for each one of the three subtests and the raw scores are translated into equivalent grade ratings, standard scores and percentile ranks. Grade ratings indicate the grade level the examinee is performing at the moment of testing.

Practical evaluation: the test provides a manual and scoring sheets. The manual, in this revised version, can be considered an improvement over earlier editions of the tests in terms of providing better organized instructions for administration and scoring. Directions are easier to follow and items are organized in a logical sequence which makes the procedures of administration less complex and confusing.

The WRAT is a fairly easy test to administer and does not require highly trained personnel.

Face validity: the test appears to have adequate face validity as a coarse screening measure in three restricted areas: spelling, word pronunciation and arithmetic computation.

Norms: The manual provides information on the normative sample. The revised WRAT was administered to school children and adults in seven states in the U.S. It is stated that no effort to obtain a representative sample was made, nor was it considered essential for proper standardization.

Norms were obtained from 5,868 subjects for Level I and from 5,933 subjects for Level II, including children and adults. Selected groups were not restricted to any economic, intellectual or racial population. No additional information is provided on the characteristics of the normative sample, excepting age distribution and the fact that all subjects in the sample were given ability tests. Half were male and half were female in the sample population.

Reliability: Reliability varies from .90 to .95, with an average of .93.

Split half reliability: high level of split half reliability are reported for all levels of age and for all three test levels ranging .94, .98.

Test-retest coefficients: internal consistency was reported as being high, the correlations ranging from .88 (for spelling) to .95 (for reading).

Validity: Concurrent Validity: Correlations between WRAT and external criteria are presented in the manual. WRAT reading test correlates with other reading tests such as the Stanford Paragraph Reading at the .80 and .90 level. WRAT and reading grades and teacher's ratings correlate approximately at .88. The WRAT Arithmetic correlates poorly with the Stanford Arithmetic at .5 and .3 for grades 5 and 6.

WRAT and intelligence test: the correlation between WRAT and WISC is poor. The correlation between WRAT and WISC Performance Scale, in all age levels of the WRAT, ranges from .68 to .52. Most of the correlations between WRAT and WISC Verbal and Full scale fluctuate around these same levels.

WRAT correlated with the California Mental Maturity IQ presents an adequate level of correlations .80, .74 and .84 for the WRAT subtests of reading, spelling and arithmetic respectively.

Content Validity: The Revised edition of the WRAT does not present data for content validity. The test does not explicate a rationale for item development. The items reflect roughly three components of school achievement as measured through ability in three skills (reading, spelling and arithmetic).

Construct Validity: The data presented as evidence for construct validity in the 1965 edition is confusing and contradictory; the authors make statements which are not always supported by the facts. The differentiation between intelligence and achievement is a difficult task. The factor analysis on the WRAT identifies seven factors influencing the subtests. Among them, the most important are: (a) a "general factor" influencing to some extent all human abilities which account for the 28% of the variance of the three scores (Reading, Spelling and Arithmetic), and (b) the Group I or Verbal factor which, as the author states, is related to personality, unrelated to intelligence. This factor would account for 30% of the variance. The remaining factors are also associated with personality factors in the levels of motivation, affect, cognition, somatic and reality. The first factor being the only one associated directly with achievement, it becomes difficult to consider the WRAT as an achievement test.

The identification of these factors underlying the test, the author suggests, may be of value, in individual cases, to provide a basis for an adequate remedial program.

General Comment: The WRAT is a reasonably easy test to give, widely used in school as a quick measure of school achievement. However, the tests do not evaluate all aspects and complexity of school achievement but a narrow range of abilities represented by reading, spelling and arithmetic. The validity of the WRAT in evaluating these three abilities is also questioned due both to the small normative sample in each range of age and the reduced number of items included in each subtest. The test results have to be complemented with IQ test results, as the author himself recommends. The main criticisms regarding the WRAT concern the normative sample and validity data. The normative sample is briefly mentioned but not well defined and described by the author. Concerning validity, the manual is rather confusing and lacking of necessary information. As far as factor analysis is concerned, the manual presents confusing statistical data and it is difficult to assess the relations between the multiplicity of factors and correlations being presented. In general, there is an excess of irrelevant data and a lack of data to support the majority of the statements made by the authors concerning norms, validity and their definition of school achievement. The fact that only a small percentage of the variance accounts for school achievement makes the test a suspicious instrument for assessing children having learning problems. There are a multiplicity of factors, which the author does not identify, interfering in test performance. The emotional aspects which, according to the authors, the test identifies, remain an obscure subject to provide useful information in assessing and programming for exceptional children. Moreover, the test is usually administered as a purely school achievement test by personnel generally unaware of the poor validity of the WRAT resulting in misdiagnosis, labelling and misplacement of children.

In conclusion, the WRAT appears as a practical, attractive, time-saving test to give to a child as a rapid assessment, although limited, of the child's abilities in reading, spelling and arithmetic. However, the examiner must be aware of the test restrictions, such as the absence of data from the normative sample, which implies a lack of reference criteria for a valid interpretation of test results. In addition, the examiner should take into consideration the absence of information on validity. The test is not measuring general school achievement but specific abilities in reading, spelling and arithmetic.

Title of Test: Canadian Test of Basic Skills (CTBS)
Revised 1976

Author: Original, E. F. Lindquist and A. N. Hieronymus.
Adapted by Ethel M. King.

Published: Thomas Nelson Co.

Purpose: To measure general areas of skills and abilities.

Type of Test: Standardized Achievement Test.

Nature of Content: The authors claim that because of the heterogeneity of curriculum this test does not measure subject areas, but general areas of skills and abilities. The areas are vocabulary, reading, comprehension, language skills, work-study skills and mathematics skills. The 1976 Revision expresses all measurement items according to the metric system. Item analysis is given for a grade equivalency in each area or for a total score. There is a grade equivalency and conversion table for percentile for the start, middle and end of year for each area.

Norms: The original standardization in 1966 was on a norm population from grades 3 to 8 of 30,000 children from 225 sample schools in Canada. It varied in location, size, type, and religious background. The 1973 standardization was with an English-speaking population using the 1966 model. The sample schools were chosen to represent all provinces, and Roman Catholic and non-Roman Catholic school. Size of school, based on the number of teachers, was also systematically sampled. Elementary schools with and without grades 7 and 8 were used. The feeder school was followed up when necessary for junior high sampling. A service is available through Thomas Nelson, publishers, to assist Boards to calculate local percentile norms and stanines. Socio-economic details and community size were not reported, although sampling was based on the 1961 Canadian Census data. The sample of schools comprised 137 units from 19 provinces, each representing one of nine school categories (a 90-cell design). Eighty percent of the schools chosen agreed to participate. Ten further schools were added to supplement the schools which opted out of the original sample.

Validity and Reliability: The CSE Elementary School Test Evaluations (1976) report that the CTBS's measurement validity has been variously rated as poor, very poor and as showing no evidence of rating for instructions, format and comprehension of different grade levels. The manual lacks information on which to evaluate test item difficulty, appropriateness, and relevance. Since the true validity for this, or any other, achievement test is the degree to which the test items probe the content actually being taught, each province and school board must examine the test content compared with the curriculum established for that system. To the extent that the CTBS attempts to sample curriculum across Canada, it will be insensitive to local or provincial emphases.

General Comments: The 1976 revision, which was somewhat more systematically normed than its predecessor, has not received any detailed reviews in the evaluation literature. However, the basic structure of the original test remains and the new edition shares some of the same strengths and weaknesses of the older version and of the Iowa Test of Basic Skills on which it was based. The following review is therefore extrapolated from criticisms made of the original tests, and considered in light of the content of the revised edition.

Birch (1972) criticized the use of a totally English-speaking school population as not reflective of Canadian norms. This criticism still holds and limits the interpretation of scores on second-language children. The original edition did not account sufficiently for curriculum and grade equivalency changes from the Iowa test, a situation partly corrected in the revised edition. However, an emphasis on capitalization, punctuation and grammar still may not be reflective of current language programs. Birch (1972) suggested that age norms established by local school boards would be beneficial. He adds that directions are clear, but the 5-hour test should be spread over a 4-day period, especially for younger children.

This test does not seem to be appropriate for children with learning disabilities. The test emphasizes reading for mathematical problem-solving, map work and comprehension areas.

Title of Test: Gates-MacGinitie Reading Tests (1965)

Author: A. Gates and W.H. MacGinitie

Publisher: Teacher's College Press, Columbia University

Purpose: Reading test evaluating vocabulary and comprehension

Type of Test: Grades 1 - 3 has vocabulary and comprehension subtests.

Population for Which Designed: Grades 1, 2, 3, 2.5 - 3, 4 - 6.

Time Required to Administer: 120 minutes

Nature of Content: The Vocabulary Test samples the child's ability to recognize or analyze isolated words.

The Comprehension Test measures the student's ability to read complete prose passages with understanding.

Norms: This series of tests replaced the Gates Primary and Advanced Primary Reading Tests and the Gates Reading Survey. Norms were based on the sample of 40,000 pupils in 38 communities which varied in size and geographical location. Pupils were of average educational level and average family income. All grades were tested in October, 1964, except for grade 1 tested in January, 1965. Students took two tests: either two forms of the same test, or one form of the test plus a second of the same form designed for the grade just above or below his own. Autumn and spring administrations were given.

Reliability: For Primary A Test for grade 1, the alternate form reliability coefficients were: Vocabulary .86 and Comprehension .83, and for grade 2: .87 and .81 respectively.

Powell (1969) states that there was no validity mentioned in the manual and content validity was not discussed. There was no proper interpretation of the scores for classroom use, and no evidence that the grade score equals the instructional reading level for the student. The speed in vocabulary and comprehension tests result in pre-frustration level or frustration scores being common. Fifteen to 20% of the pupils could not complete the test in the allotted time. The test is of limited value if information is needed about specific reading subskills. It is usable for comparison to battery tests.

The Center for the Study of Evaluation (1976) evaluated the Gates-MacGinitie as generally poor or with no detail at all in all these areas: measurement validity, examiner appropriateness, normed technical excellence and administrative validity.

Burke (1972) criticizes the lack of a rational view of the reading process in the manuals. Vocabulary involves an overly minute and detailed word attack situation where the child matches very fine graphic and phonemic discriminations between words with minimal contrasts. The examinee is tricked or confused as the section continues, and at the end of the first-grade and in the second and third-grades, less familiar stimulus words are used. Two examples are: a picture of a comb and the words comb and come, and medal and meddle. At levels 4-9 there is confusion, and a time loss could result in choosing the words closest in meaning to an isolated word.

The speed and accuracy subtest for grade 2.5 and higher measures how rapidly students can read with understanding material the test authors call "of relatively uniform difficulty." Burke (1972) criticizes this as being irrelevant to the reading process, and simply a measure of word association time. Such a speed test can cause anxiety; when the speed constraints were removed, a sample of grade 6 children registered significant gains on this subtest.

In the comprehension subtest, at levels A and B the student selects one of four pictures to match a sentence or paragraph, and at level C, there are multiple choice questions. At levels D and E there are cloze-like procedures. Burke (1972) claims that since all four pictures can involve a major word, it is vocabulary and not comprehension which is measured. At level C, 21 of 48 items can be answered without reading a paragraph. This test calls for direct recall of a common situation, and therefore vocabulary more than comprehension is also being tapped. Thus, the skills and strategies required for success may have no direct relationship to successful reading in the classroom context. Comprehension, the supposedly more important area of this subtest, is poorly measured.

Berg (1972) reviewed the Readiness Skills Test of the Gates-MacGinitie Reading Tests for K-1 with 9 scores. He believed that the 120-minute test is excessive in length, containing more subtests than the closest comparable test. The stimulus items are clearly defined, the pictures carefully detailed, the words in primer type are spaced well for ease in discrimination. The correlation for grade 1 between total readiness scores and vocabulary and comprehension scores of the reading test were .60 and .59 respectively for tests administered in October and May. A large amount of variance is not accounted for between predictive measures and their criteria. Normative data seems adequate within the limits of the variance. This test can measure a child's ability in a limited number of areas according to the test authors. When used as a diagnostic tool it should be supplemented by formal and informal evaluation to extend the limited range of skills being tapped.

Dykstra (1972) claims that the test authors do not indicate whether the test is to identify pupils ready for normal reading instructions or to diagnose strengths and weaknesses in pupils' readiness profiles. In Dykstra's view, the subtests performance should be more important than the total score for diagnostic use. Dykstra notes the following weaknesses:

1) The norm tables indicate little difference in performances between the end of Kindergarten and the beginning of grade 1;

2) The manual fails to clarify whether the test is designed to predict reading readiness specifically or school readiness in general.

3) No suggestions on how to interpret any given score are given.

Dykstra also questions the artificial nature of auditory blending test.

Van Roekel (1972) reports that the only changes in the new edition of the test are in format and structure. Eller (1965) criticized the word recognition subtest of both the earlier and revised version. Eller notes that the sight vocabulary test is restricted to words easily defined by pictures. The result is that 90% of the revised test's items required noun identification only. Unfortunately, short function words ("these," "from," "was") account for a large proportion of word recognition errors in grades 1 and 2 and are not accounted for in the interpretation of results. However, Eller praises the organization of the manuals and the procedures for standardization.

In summary, the Gates-MacGinitie, a long-standing member of the reading test arsenal, seems to be a reliable, well-standardized test for those limited skills which it taps. It is probably not a test of reading in the broad sense, but a test of vocabulary and word recognition skills, and therefore of limited predictive validity for both normal and reading problem children.

Title of Test: The Peabody Individual Achievement Test (PIAT) 1970

Author: Dunn and Markwardt

Publisher: American Guidance Services

Purpose: To provide a wide range of quick screening tests for individual achievement

Type of Test: Achievement (individual)

Population for Which Designed: Kindergarten to Grade 12

Time Required to Administer: 30 to 40 minutes

Nature of Content: There are six scores for individual achievement: Math, Reading Recognition, Reading Comprehension, Spelling, General Information and the total.

Norms: The actual standardization sample was children in regular classrooms in public day schools in nine regions in the U.S. The sample included urban and suburban children from each region. There were 200 subjects for each of the 13 grade levels, half male, half female. There were 84.4% white, 11.39% Negro, and 4.3% other.

Content Validity: The tests show a .57 overall median test-subtest correlation. The content was specifically selected to minimize bias towards any given curriculum approach.

Reliability: Test-retest reliability coefficients for the PIAT range from a low of .42 to a high of .94, with an overall median reliability coefficient of .78. The standard errors of measurement are somewhat high, ranging from a median of 3.06 for Reading Recognition to 6.51 for Reading Comprehension, with a total test median of 12.05.

Construct Validity: French (1972) claims that the PIAT locates the general level of school achievement and in addition a more comprehensive test should be used. Improper use of a profile may result from a lack of a psychometric skill in the tester, a factor which the manual ignores when it claims that "non-professionals" may administer the test. French calls the total score more reliable than the subtest scores. PIAT's correlation to PPVT IQ's range from .53 to .79 and the median is .68. The total test's reliability after one month is a median of .78. Lyman (1972), however, notes that the demonstrated validity and test-retest reliability is low and requires further research.

General Comment:

Lyman (1972) considers the content good, and the items appropriate for each of the five subtests. The test begins with the assessment of mathematics skills on tasks that do not require reading or writing. Bannantyne (1974) considers this non-requirement and the pictorial and spatial basis of the comprehension tasks offer possibilities for testing learning disabled children who have difficulties in expressive language. On the other hand, 40% of the total test scores consist of spelling and reading pronunciation.

Both Bannantyne and Lyman consider the format of the test to be superior and the quality of materials excellent, although the record booklet is somewhat crowded.

Proger (1972) approves of the field testing, item difficulty and item discrimination used. He challenges the test authors who call it a test of power not speed, especially when it is being used with exceptional children. Proger suggests breaking the rules and allowing sufficient time to finish the test, especially with children suspected of learning difficulties. For example, he disapproves of the rule that if after 3 minutes a child cannot score one correct item in the training section, the subtest should be terminated and the child receive a raw score of 0. Proger also criticizes the use of a mathematics subtest to guide the starting points for a reading recognition subtest. In his view, it is unfair that disabled readers be given an excessively high basal level of verbal functioning due to math scores.

He suggests that testers should use their clinical judgement in choosing the starting points for each subtest.

Proger considers PIAT's validation to be weak, but the content validity sound and concurrent validity good. The subtests and the total score correlated well with the Peabody Picture Vocabulary Test. There is superior standardization and construction. The subtests are adequate, the manual documentation is excellent, although the validity studies and the basal-ceiling procedure can be criticized. Proger views this test as a sound challenge to the WRAT.

Lowrance (1977) used the PIAT to determine the stability of the WISC-R in a learning disabled population of fourth-graders. His goal was to determine the best predictors of reading comprehension for the WISC-R, to distinguish between mildly learning disabled and severely learning disabled children. He found that reading could be predicted, the two groups of children were objectively discriminated and that both tests were sensitive to socio-economic levels.

Predictive Validity of Groups of Tests, Including Standardized Achievement Tests, for Disabled Learners

a) Studies using the Wide Range Achievement Test (WRAT)

Camp (1973) examined the power of several tests to predict the learning levels and rates of severely disabled readers. He selected a test of auditory-visual integration by Kahn and Birch (KB, 1968) and a test of visual-spatial perception by Sterrit and Rudnick (1971), together with the WRAT, Bender-Gestalt and Raven's Matrices.

Subjects were 69 children with severe dyslexia (2-5 years below grade level), aged between 9 and 13 years, and with an IQ greater than 75. Those children in the third and fourth grade were eligible only if their reading score was at least two years below actual grade level. Learning rates were also assessed for 42 of the children.

Significant correlations were found among the psychometric tests, but not between the tests and learning rates. Children achieving above and below expectancy in their learning programs differed significantly in learning rate and amount of gain over a six-month period, but not on any of the psychometric tests.

The authors conclude, as did Cohen (1979), that the presence of perceptual handicaps does not predict response to learning in a remedial program. Learning rate is highly predictive of achievement, but that standardized measures which purport to tap various skills associated with learning potential have very little predictive validity, and contribute little in explaining individual differences in learning.

By contrast, Ferinden, Jacobson and Linden (1970) used the WRAT, the Evanston Early Identification Scale (EEIS), the Bender-Gestalt Visual Motor Test and the Metropolitan Reading Readiness Test Form R to form a test battery to diagnose potential learning disabilities at the kindergarten level. Sixty-seven children from eight classes with a high percentage of culturally deprived children and two classes from mostly middle-class backgrounds were tested and clinically judged by a team composed of a school psychologist and two learning disability specialists. They found a 93% accuracy from the WRAT results in identifying those children who were later unsuccessful in learning to read at the first-grade level. The authors suggest that kindergarten teachers could be trained to administer the WRAT and the EEIS, and that this might provide "almost 90% accuracy in screening for potential learning disabilities" as a result.

b) WRAT with Keymath

Kratochwill and Demuth (1976) examined the predictive validity of the Keymath Diagnostic Arithmetic Test (KM) (Connolly, Natchman & Pritchett, 1972) and the WRAT's (Jastack, Bijou & Jastak, 1975) Arithmetic subtest. Thirty-seven children from a rural Wisconsin community in a mainstreamed program with supplementary instruction participated in the study. Their standard achievement test scores were one year or more below grade level in mathematics and they were of average intelligence. Trained paraprofessionals individually administered the KM and the WRAT's Arithmetic subtest, while the Metropolitan Achievement Test, Primary I Form F was given by the classroom teacher as a criterion achievement test.

The KM related weakly to the WRAT ($r = .32$, $p < .05$), suggesting that the two tests measure different areas of mathematics performance. There were insignificant correlations between the WRAT and the Metropolitan ($r = .17$, $p > .05$), but the KM correlated .63 with the Metropolitan ($p < .001$). The two correlations differed significantly ($p < .01$).

The KM was found to be a moderate predictor of school achievement, accounting for about 40% of the variance in Metropolitan scores. The limited range of items on the WRAT subtest may have depressed the correlations, as well as the restricted age range and the poor mathematics performance of the sample. As a diagnostic test for learning disabled children, the KM is attractive due to the broad range and diversity of item content, lack of reading and writing requirements and good predictive relationships with school achievement. However, more research on its reliability with this population is needed.

c) The Peabody Individual Achievement Test (PIAT)

The PIAT is recommended by Bray and Estes (1975) for learning disabled children because the child with fine motor co-ordination difficulties may point to an answer, arithmetic word problems do not require reading, and the overall format is less formal than other tests. A "substantial concurrent validity" is reported, although the PIAT was standardized on a normal population.

Wilson and Spangler (1974) noted the speed and ease of the PIAT's administration, the few potentially distracting test materials used, the lack of required written responses, and the minimal requirement for lengthy verbal responses. Eighty-three children and adolescents referred because of learning disabilities were tested using the PIAT and one or more standard measures of intelligence, the WISC or the Stanford-Binet, and in 57 cases the Peabody Picture Vocabulary Test as well. When partialled for age, a moderate positive correlation was found between the PIAT and all three intelligence scales. Internal consistency proved high for the PIAT subtests, supporting the reliability coefficients reported in the Manual for a normal population, and standard errors of measurements for the five subtests were relatively low, ranging from 2.20 (Reading Comprehension) to 2.88 (Mathematics).

Harmer and Williams (1978) compared the WRAT (1976) and the PIAT's (1970) three corresponding subtests. Trained clinicians at a university-affiliated Learning Disabilities Center tested 66 children on the Word Recognition subtests, 62 on Spelling and 66 on Math. Moderate to strong positive correlations between subtests were obtained (Reading, $r = .87$, $p \leq 0.5$; Spelling, $r = .73$, $p \leq .05$; Math, $r = .64$, $p \leq .05$). There were significant differences in math scores at grade 4 and beyond. The spelling and reading subtests indicated a much greater overall similarity. Some math scores indicated differences as great as 6 years in grade equivalence after grade 4. The math subtests are not interchangeable.

The WRAT (1976) is more appropriate for high school students, while the PIAT is most sensitive at lower levels. Besides content differences, there are procedural differences in administration, test format, and method of scoring. The PIAT is untimed, the WRAT has a time limit of 10 minutes for subtests, but it is usually ample. There is more continuous interaction during the PIAT between the examiner and testee. The PIAT's multiple choice format tends to be less accurate than the WRAT, permitting guessing, with in some cases a 50% chance of being correct. The PIAT's ceiling for grade 12 is poor.

Wetter and French (1973) administered the PIAT and the WRAT to 30 children in a learning disability clinic in order to validate the PIAT against the WRAT as a criterion measure. Their achievement levels were two years below CA grade-level expectancy. The order of administration of the two tests was reversed for half of the group. The age range was 5.5 to 10 years and the WISC (Wechsler, 1949) Full Scale IQ Index ranged from 81 to 140, with a mean and standard deviation of 103.9 and 13.2 respectively.

All coefficients of correlation were significant at the 0.1 level. Correlations were .79 between PIAT Mathematics and WRAT Arithmetic, .96 between PIAT Reading Recognition and WRAT Reading, .96 between PIAT Reading Comprehension and WRAT Reading, and .86 between PIAT Spelling and WRAT Spelling. The authors suggest that the tests measure similar areas of achievement, and thus are interchangeable in assessment of learning disabled children. However, with this small a sample, and large a range of age and IQ, the reported correlations are suspect. Further doubts about the correlation of WRAT and PIAT are provided by Sittingdon (1970), who states that, with the possible exception of the reading subtests, they differ in content and approach. As Wetter and French remark, further validation of the PIAT is needed.

Soethe (1972) studied the concurrent validity of the PIAT with the WRAT and the Wechsler Intelligence Scale for Children (WISC) with 40 children between 7.9 and 16.1 years of age (26 male, 14 female), labelled as "normal", "reading disabled", or "mentally retarded." "Normals" had a WISC Full Scale IQ (FSIQ) of >86, the "mentally retarded" had a WISC FSIQ of <85, and the "reading disabled" a WISC FSIQ of <86. The results showed a substantial correlation between the PIAT and the WRAT subtests in Reading Recognition and Comprehension, despite test differences. Moderate correlations were found between the PIAT and WRAT Spelling and Arithmetic subtests. The PIAT Math correlated more highly with the WISC Arithmetic than with the WRAT. The information subtests had moderately high correlations between the PIAT and WRAT. Pearson correlations between the PIAT and WRAT subtest scores ranged from .44 to .87 for the mentally retarded, and from .37 to .92 for the reading disabled. Correlations of the PIAT with the WISC FSIQ ranged from a low of .22 for the mentally retarded group to a high of .80 for the normal group. Soethe concluded that the PIAT could be used in conjunction with the WRAT and WISC for differential diagnosis of learning disabilities, but used alone, its predictive power is poor.

Bray and Estes (1975) evaluated the concurrent validity of the PIAT by using the California Achievement Test (CAT), WRAT, and Teaching Ratings of academic performance as criterion measures. The subjects were 45 Caucasian children (37 boys and 8 girls), classified as learning disabled, from a middle-class socio-economic group. The children were students at a Research and Evaluation Center for Learning. Their ages ranged from 7 years, 1 month to 12 years, 9 months and their WISC FSIQ scores one year prior to the study ranged from 72 to 125. A school psychologist administered the tests and qualified special education teachers rated the children's achievement level in grade equivalence after eight months of instruction.

The PIAT, CAT, WRAT and Teacher Ratings (TR) resulted in moderate to high correlation coefficients in all areas. Comparable subtests mean grade placement scores were generally equivalent. Exceptions were the PIAT Reading Recognition and the WRAT Reading scores, which were approximately 0.5 years lower than those obtained on the other two measures. This may have been the result of requiring an oral response on the PIAT and WRAT subtests, but a visual-motor response on the CAT.

The lowest correlations in reading were between the CAT Reading Comprehension with both PIAT Reading scores and the CAT Reading Vocabulary. In mathematics, the general pattern revealed moderate correlations between CAT Arithmetic Reasoning and the remainder of the Arithmetic scores, with the exception of the CAT Total Arithmetic. The PIAT Mathematics and Teacher Rating were the lowest correlation. All correlations in Spelling showed highly dependable relationships. The Total Test Battery scores from the PIAT and CAT showed a correlation of >.77, indicating a fairly high dependable relationship.

Bray and Estes (1975) see the following components of the PIAT as valid for diagnosis of learning disabilities: (a) the auditory component on the Mathematics subtest; (b) oral identification of letters and words in the Reading Recognition subtest; (c) silent reading followed by a choice of four pictures in Reading Comprehension; (d) a choice of the correctly spelled word, rather than a dictation in the Spelling subtest; and (e) a subtest measuring general funds of knowledge. They note limitations regarding validity, reliability and the lack of sufficient differentiation of raw scores after fourth grade.

(d) The PIAT and California Achievement Test (CAT)

Watson (1972) found high correlations in comparing the PIAT to selected CAT subtests for 94 educable mentally retarded (EMR) and 116 educationally handicapped (EH) students. For the educationally handicapped group, Reading Recognition correlated with Reading Vocabulary (>.92), with Reading Comprehension (>.84) and with Mathematics and Arithmetic Fundamentals (.79).

(e) The PIAT, CAT and Metropolitan Achievement Test (MAT)

Bray and Estes (1975) correlated the PIAT and the MAT on a sample of 222 fourth-grade children in Dallas from a broad socio-economic background. There were 107 Learning Disabled (LD) (63 males and 44 females) and 115 Non-Learning Disabled (NLD) (56 males and 59 females) subjects.

Moderate to high correlations were demonstrated. For the LD sample, correlation coefficients ranged from .45 to .86. Marked degrees of relationship were observed between the PIAT and MAT reading scores. Moderate correlations were obtained between the two spelling scores, and between the PIAT Mathematics and CAT Arithmetic scores.

Obtained correlations indicate substantial to marked degrees of relationship, with four exceptions. For LD females and NLD males, only moderate instead of high correlations were observed. The same decrease in magnitude was obtained for the spelling subtests for both LD and NLD females

From the results of their two studies, Bray and Estes conclude the following:

1. Substantial validity is demonstrated in support of the PIAT when using the WRAT, the CAT, or the MAT as criterion measures.

2. The highest correlations were observed between the PIAT and WRAT as expected, since both are screening instruments with relatively few items at any one age or grade level.

3. There were sufficiently high correlations between the PIAT and both the CAT and MAT related subtests.

4. Although correlations between PIAT Mathematics and MAT Arithmetic scores were of lesser magnitude than noted for the corresponding CAT and WRAT subtests, moderate degrees of concomitance were still reflected.

5. With the exception of Mathematics, correlations appear relatively high when computed for the NLD sample or for both the LD and NLD groups classified by sex.

Bray and Estes see the PIAT as an index for predicting academic potential in learning disabled children, when used in conjunction with a variety of other tests. The format, interest and motivational factors may recommend its use in clinical and diagnostic settings. However, as an individual diagnostic test of learning disabilities, further research is needed, especially in validating the PIAT as a predictor for learning potential and classroom performance.

(f) Metropolitan Achievement Test

Belmont and Belmont (1978) studied the MAT reading test scores for a graduating class of a New York City public elementary school (131 boys and 112 girls) with approximately 80% from white, middle and lower middle-class families. They examined individual patterns of performance to determine the degree of consistency or variability in reading achievement across the grades in terms of six functional levels from poorest to the best. The correlational data suggested that the entire group was more or less consistent in test performance over time; however, when examined in terms of the consistency of each child's performance over time, there were large amounts of variation in individual performance.

Individual variability across the grades showed that: (a) only a small number of the children's reading scores were consistent as they progressed through the 12 grades, and those that were consistent were predominantly the better readers; (b) a majority showed functionally significant changes in reading level; and (c) the vast majority of children varied by as much as two or three grade levels of their own mean scores.

The correlational analysis of reading scores obtained over the grades showed a large positive relationship between earlier and later reading performance, attesting to the reliability of the MAT measures on traditional correlational criteria. Yet 24% of children were classified as among the poorest readers in one grade or another, but only a small proportion of these were consistently poor.

Belmont and Belmont's findings are consistent with those of Traxler (1960) and Clarke (1972), showing that individual achievement is more variable than stable, and that standardized tests scores, based on norms, are not reliable for predicting the progress of a given child. This finding contrasts the frequently held view (e.g., Bloom, 1964) that reading achievement is a stable and consistently developing trait, similar in essence to IQ. On this point, the International Reading Association now has an official policy of discouraging the use of standardized reading tests, stressing the need for the diagnosis of reading as a multi-faceted, complex skill.

These studies, which provide evidence of periods of slow growth and of spurts in normal children's reading development, suggest that early identification may be hazardous. The use of the MAT and other standardized instruments may not be valid for providing long-term predictions of learning potential or the need for remedial intervention.

The Learning Disabled Pupil in Ontario: Report of Interviews Conducted on a Sample of Ten Ontario School Boards

This study was conducted between May 25th, 1979 and June 9th, 1979. Structured interviews were conducted in ten Ontario School Boards. One interviewer visited the superintendent of special education, his designate or counterpart in each of the ten boards. The discussions centered around the policies and practices used by boards in identifying, assessing, placing and programming learning disabled students. The board representatives were informed by telephone of the nature and content of the matters to be discussed prior to the actual visit. It was believed that by using this particular format, the information derived from board representatives would be more accurate. This proved to be a useful approach since all the interviewees had prepared a synopsis of current data in readiness for the interview, and this was in all cases handed to the interviewer.

Two criteria were used in attempting to select the ten boards which were visited. The first criterion was to select boards representing a range of locations and geographical areas. The second criterion was to sample boards representing proportionately high, moderate and low enrolments in both general and specific learning disability classes.

The ten boards interviewed were selected from the data available in a survey of Special Education in Ontario (Keeton, 1978). The following is a tabular representation of Keeton's data. It lists, board by board, the number of segregated special classes designed for slow learners and for specific learning disabled children. In the survey, boards were requested to list the number of segregated classes for varying categories of special needs students, one of which was "slow learners/GLD moderate", and the second one was "specific learning disability". On this table is also indicated those boards in which interviews were conducted.

The name of the ten boards selected for interview, and indeed all the boards who responded to the survey, will not be identified in the Report. Instead, we present our general conclusions and impressions from the sample interviewed.

Table 1

Boards Identified with Segregated Learning Disability Classes and Persons who do Diagnostic Testing*

Computer #	Number of Classes Slow Learner	Specific Learning Disability	Total	Diagnostician**
001(I)	70	10	80	Educational Assessment Department
002	35	4	39	Psychometrician
005(I)	201	20	221	Psychologist/Psychometrician/ Resource Teacher
006	17	38	55	Psychologist
009(I)	61	20	81	Psychological Services
010	24	-	24	Psychologist
011(I)	16	5	21	Special Education Consultant
012	5	-	5	Student Services Committee
013	18	2.5	20.5	Psychometrician
014	8	1	9	Psychometrician/Resource Teacher
015	9	-	9	Psychometrician
018	9	2	11	Psychometrician
019(I)	46	30	76	Psychometrician/Psychologist
021	32	5	37	Special Education Co-Ordinator
023(I)	45	23	67	Psychologist

* Data based on 80 responses to a questionnaire study by Keeton (1979).

** Diagnostician - refers to person or persons doing intelligence testing. Designation of this term to these personnel is based on the reliance of intelligence test scores as major discriminating measure for the identification of learning disability.

(I) Selected for interview.

Table 1 (cont'd)

Computer #	Number of Classes Slow Learner	 Specific Learning Disability	 Total	Diagnostician
024	4	11	15	Student Services Team
029	8	5	13	Teacher Diagnostician
030	15	5	20	Special Education Consultant
032	6	1	7	Psychologist
035	8	0	8	Special Education Co-Ordinator
037	45	2	47	Psychologist/Consultant
040	3	0	3	Special Education Teacher
042	19	5	24	Psychometrician
045	3	2	5	Principal/Special Education Teacher
047	12	3	15	Not identified.
048	5	0	5	Special Education Co-Ordinator
051	-	3	3	Psychometrician
052(I)	14	11	25	Diagnostic Prescriptive Teacher/ Psychologist
053	3	0	3	Child Development Centre
054	9	0	9	Teacher
055	6.5	4	11.5	Student Services
056	2	-	2	Teacher
058	19	6	25	Technical Diagnostician
059	19.5	2	21.5	Clinical Services Department
060	0	22.5	22.5	Outside Consultants
063	15	2	17	Student Services at Special Education Department
064	13	6	19	Special Education Services/ Outside Clinics
065	9	0	9	Special Education Department

Table 1 (cont'd)

Computer #	Number of Classes Slow Learner	 Specific Learning Disability	 Total	Diagnostician
067	4	1	5	Psychometrician/ Special Education Consultant
068	10	1	11	Psychometrician/Special Education Consultant
069	17	3	20	Psychologist
070	8	5	13	Psychological Services
071	5	-	5	Consultant/Teacher
072	13	7	20	Psychologist
074	18	8	26	Psychological Services/ Psychologist
075	15	-	15	Psychometrician/ Teacher Diagnostician
076	-	1	1	Teacher Diagnostician/Psychologist
077	10	1	11	Psychologist
078	35	6	41	Not available
079(I)	148	46	194	Psychological Consultant/ Psychologist
080	8	3	11	Special Education Consultant
081	9	5	14	--

The Ministry of Education has divided the Province of Ontario into geographical regions. A review of learning disability assessment policies and procedures were elicited from five boards in four of the nine regions. These regions are located in the (a) Central Ontario Region, (b) Midwestern Region, (c) Niagara Region, and (d) Western Ontario Region. Collectively, these regions represent 28,369 elementary Special Education students of the provincial total of 35,159, and 965,783 of the 1,329,396 provincial elementary school population.

The boards were chosen as representing the highest identified population of learning disabled students in either a segregated, integrated or partially integrated program. This information was based on data derived from "A Survey of Special Education in Ontario", conducted by Dr. A. Keeton in August 1977, and from Ministry data.

Special education at the secondary level is not as well defined categorically as at the elementary level. Although the statistics reported are only for the elementary level, it was found that in many cases the administrative decisions and policies affecting secondary special education are the responsibilities of the administrative staff also responsible for the elementary level. Thus the findings reported may be tentatively accepted as basic policy for the secondary learning disability population. Caution must be used however, if definitive statements are to be made about the secondary learning disability population on data derived from this section of the report.

This field report is divided into four sections. The first deals with definitions, terminology, and evaluative criteria used by boards regarding the assessment of learning disabilities. The second reports on the actual assessment procedures. Thirdly, the instruments used are reported. The final section includes concluding remarks.

Definitions, Terminology and Evaluative Criteria

Definitions. The type of definitions and terminology used by Ontario School Boards reflect the state of the field as indicated elsewhere in this report. There is no general consensus as to what constitutes a learning disability. Although boards use the Ministry of Education (Province of Ontario) definition, some use it mainly as a branching point, and amplify or emphasize to varying degrees the principles outlined in that definition. In many cases boards bypass the definitional issue and their working definitions can be inferred only from their procedures.

Three definitional components are recognizable: (a) an intelligence component, (b) an exclusion principle, and (c) etiological factors. The degree of emphasis placed on any of these components varies remarkably.

The majority of boards interviewed felt that the intellectual level of the student played a very important role in deciding if the student was to be considered learning disabled as opposed to any other category of special needs learner. Some boards omit the intelligence component entirely. They believe that the intelligence of a student is secondary to the immediate educational needs of that student and the instructional capabilities of the teacher. By far, boards adhering to this type of philosophy are in the minority.

The exclusion principle was at work in one form or another in all the board philosophies reviewed. All boards excluded the mentally retarded from learning disability programs. The remaining groups of disabled learners were included or excluded depending on the philosophy of the board. It was observed that the deaf, blind and partially sighted were included or excluded in direct relation to the policy of the individual boards regarding integration of handicapped children into mainstream education.
The remaining groups of disabled learners were included or excluded depending on the learning disability definitional philosophy of the board pertaining to these groups.

The gamut of learning disability classes ranged from specific learning disability classes to behavioral learning disability classes and included learning disability and speech and language learning disability. Students were included or excluded from these types of learning disability programs based on the admission criteria set down for each of these types of programs. In some cases, particularly the general learning disability programs, one of the criteria tended to be that the student had been unable to benefit or was thought unable to benefit from any other type of special education program which the board offered. No boards considered cultural factors in the exclusionary process.

Etiological factors entered into definitional components in a definitive manner only in a few boards. Neurological dysfunction of the CNS was not generally considered a major contributory factor to a learning disability condition, with only one of the boards interviewed indicating that presence or absence of neurological dysfunction was a major consideration.

All boards considered deficits in perceptual and perceptual motor functioning as major etiological factors affecting students with learning disabilities. It was difficult to ascertain if these boards distinguished between CNS dysfunction and perceptual and perceptual motor deficiencies. The impression was gained in most interviews that the majority of boards equated dysfunctions in perceptual processes and neurological processes. Environmental factors (e.g., food additives, nutrition) were not mentioned as etiological factors.

Terminology. A variety of terminologies exists within the Ontario school system, with reference to students considered to have a learning disability. This may result in a confused state of affairs if one attempts to list the terminologies used and accompany them with definitions or explanatory statements. The same term may have different connotations in different boards. A most notable example is the term "general learning disability". In many boards the use of this term indicated the severity of the learning disability condition. These boards reported the term as meaning a student who was performing adequately in overall academic areas, but had specific problems in one subject area; in other words, students who may be having problems in areas related to reading and spelling but were successful in mathematically related subjects. The other view is held by those who maintain that a general learning disability is demonstrative of a student experiencing global difficulties in all dimensions of school achievement.

The reverse corollary of this is also true. Some boards consider as learning disabled those students who demonstrate an inability to cope in any academic strata because of a specific learning disability condition which interferes with their cognitive ability. The other extreme of this concept results in boards defining a specific learning disability student as one who performs well in all areas except for a specific subject where he is unable to gain academically.

The discrepancy in terminology is indicated by labels attached to learning disability classes. At least nine such descriptive labels are in use by Ontario boards. These include learning disability classes, general learning disability classes, specific learning disability classes, occupational classes, comprehensive classes, behaviorally learning disabled classes, language disability classes, speech disability classes, and classes for the perceptually handicapped. One final notation regarding terminology is possibly in order here. In many boards, it was found that the term "perceptually handicapped" was used interchangeably with "learning disabled."

A few boards attempted to distinguish philosophically between a general learning disability and a specific learning disability. They reported that in practice, it is a very difficult process to find that fine dividing line. In the light of this fact, a large percentage of boards contacted preferred to use the term "perceptually handicapped" rather than attempt to distinguish between the general and the specific learning disabled. In fact students who traditionally may be referred to as remedial education students are included in some Resource Teachers' learning disability programs.

Reflecting the practical problems involved in operationalizing the basic definitions in the field is the variety of ways in which personnel in the field are employed. Resource teachers are used in some boards as in-school consultants, in others as diagnosticians and elsewhere as special education teachers. A special education teacher may find himself or herself having any of the above mentioned responsibilities, or be a teacher of a segregated special education classroom.

The labels attached to varying degrees of learning handicaps have meaning only within the parameters of a particular board. For example, a slow learner was considered by one board as being an educable mentally retarded student who would not be eligible for learning disability classes. In another board, the slow learner was included in learning disability classes, and was not considered an educable mentally retarded student, but one of normal intelligence having overall academic problems. A third board considered a slow learner as a general learning disabled student, and another included the slow learners among the remedial students. Perhaps it might be correctly stated, with some caution, that this situation of ambiguity in operationalizing procedures based on the terminology used within the Ontario school system is also prevalent for the terminology in use to identify the majority of special education students.

Evaluative Criteria. It might seem reasonable to assume that within any assessment procedure the types of evaluative criteria in use would determine the nature of the evaluative instruments used subsequently. It is therefore important that an examination or identification of the criteria upon which an assessment is based be made. These criteria are recognizable through a review of the specific admission criteria set out by boards. It is notable that these criteria tend to be in direct relation to definitional concepts adopted by boards.

The most prevalent evaluative criterion in use is a psychological assessment. The psychological component was reported as being, if not the most important, at least a very important part of the diagnostic process in all boards. Thus, a deficit, as identified in a psychological assessment, is a very important criteria for inclusion in learning disability programs. Average or above average intelligence was the second most consistent criterion for inclusion in learning disability programs, and the third criterion was an educational lag of two academic years. In some cases a medical examination was used but was not mandatory. All the boards contacted maintained that the student must be unable to benefit from, or had been unable to benefit from, other special education programs.

The Assessment Process

The actual assessment takes place on two levels in most boards. The first may be called the inschool level, and the second, the psychological assessment level. Assessment at the inschool level consists of standardized or informal testing done by principals, regular classroom, special education, or resource teachers. To this point the student has demonstrated an inability to benefit from regular instruction. If he is still unable to benefit from instruction, and shows no progress after being exposed to programming resulting from an inschool assessment, he is then referred to the psychological assessment level.

Many boards have their own psychological assessment centres, or student services departments. By and large all boards interviewed made such services available to their students. If boards had not established a centre of their own, they bought services from other boards or from private professional assessment centres. Within the boards contacted, parental permission was always sought for evaluation at this level. No testing is done at this level by the principal or classroom teacher.

The major learning disability testing takes place at this level with tests being administered by psychologists, psychometricians, resource teachers, special education consultants, speech therapists or language therapists. These diagnostic clinicians quite often team their testing results and reports, have discussions and exchange ideas. Their final report then becomes the assessment report. The school nurse and sometimes the social worker may be included to provide observational information about home conditions or school behavior but this is not universal. Assessment at this level may sometimes be done by only one of the aforementioned clinical diagnosticians, who singly submits an assessment report. There is no consistency in the professional background of such persons.

The assessment report goes back to the principal of the individual school, who initially requested the assessment, or to the Placement and Review Committee of the board. Some boards attempt to bridge the distinctive break between the inschool assessment and the psychological assessment in one of two ways. The first allows the principal and sometimes the special education or regular classroom teacher to sit in on Placement and Review Committee meetings. The second method provides for the special education consultant to act as a liaison between the school and the assessment centre.

Instruments Used

Irrespective of the differences in definitional components, terminology, or evaluative criteria, the measures used for assessing a learning disability are similar across boards. The only major deviation is with respect to boards who dissociate themselves from the intelligence component of the learning disability definition and thus omit intelligence testing.

Many boards indicate that the choice of tests for a particular student varies and depends on the indicators provided by key test results when once the testing has begun. Based on this information, providing a list of tests used by boards would not do justice to the present discussion. It is perhaps more useful to identify the major categories of tests used, relating them to the two assessment levels previously described and to provide a few examples. Four major categories of tests are currently in use. These are (a) intelligence tests, (b) gross and fine psychomotor tests, (c) auditory and visual perception tests, and (d) educational diagnostic achievement tests.

At the inschool level, most of the testing is done using the educational testing category. The main tests used here are: the Standard Diagnostic Tests (Reading and Arithmetic), Durrell Diagnostic Reading, C.T.B.S., and other standardized diagnostic and achievement tests. Individual teacher-made informal tests and checklists are frequently used at this level.

Intelligence, psychomotor, and perceptual testing takes place at the psychological assessment level. Intelligence testing leans heavily on the use of the WISC, WISC-R, and WPSSI, with many boards reporting that these tests are the major determinants of a learning disability condition. Boards often suggested that a discrepancy between verbal and performance scores, derived from the WISC or WISC-R, indicated major areas in which further diagnostic testing was necessary, and was a major diagnostic criteria. The Slingerland Screening Tests, Illinois Test of Psycholinguistic Ability, Peabody Picture Vocabulary Test, and Wepman Auditory Discrimination Test overlapped into the areas of intelligence testing and perceptual testing, where a language deficit was included as a perceptual deficit. The most popular tests used at the psychomotor level were the Bender Visual Motor Gestalt Test, and the Frostig Developmental Test of Visual Perception.

Summary

A review of the infield assessment policies and procedures in use by a sample of Ontario school boards indicate that the assessment procedures and policies for learning disabilities are not uniform throughout the province. The type of assessment procedures adopted is dictated by the theoretical framework within which each board functions.

The definitions of learning disability adopted and the terminology used to describe the definitional components determine what use is made of information derived from an assessment. This assessment occurs at two levels: an inschool level and a psychological assessment level. In determining the presence of a learning disability, the psychological assessment level takes precedence over the inschool level.

It should be noted that there does not exist within the province a distinctive learning disability assessment procedure. The diagnosis of a learning disability is an outgrowth of the regular special education assessment process, with an intelligence test component, in the majority of cases, being a major deciding factor. There appears to be no clear-cut demarcation between a specific learning disability, a general learning disability, and retarded school achievement. The severity of school achievement retardation frequently determines the category of learning into which a particular student is placed.

It appears that learning disability classes and programs are set up along a continuum, based on academic success, or, perhaps better stated, the lack of academic success in school: the general learning disability at one end of the continuum, and the student demonstrating the most severe academic retardation being termed "specific learning disability" at the other end. Segregated learning disability placement is a last resort. The label is attached with the underlying belief that if the educational strategies previously prescribed at the inschool level did not work, then the problem has to be psychological. This then warrants the administering of tests within those categories previously described, namely, intelligence tests, psychomotor tests, and perceptual tests.

The major emphasis of the assessment procedure is the examination of the child for deficits. Little attention is paid to the examination of the educational situation in which the student was asked to function. Educational input on the assessment team at the psychological level is frequently lacking. Quite often it was reported that the assessment team takes into consideration an educational assessment prepared at the inschool level. This component of the assessment process is thus being derived from the very agencies who were unable to provide effective programming initially. In cases where the educational assessment is not performed at the inschool level, adequate coverage of this area is dependent upon the clinical experience of the assessors, who may have gained this expertise from assessment situations of a clinical or other non-educational variety, or in which the educational component was similar to the present situation.

Thus in the final analysis, two important considerations evolve through a study of learning disability assessment procedures and philosophies of a representative sample of Ontario school boards.

1. Since learning disability assessment is the final by-product of special education assessment, there exists a need to consider whether or not the learning disability category provides an additional label encompassing learners whom the educational system has failed to accommodate, especially since the assessment procedure focusses on the ability of the student and not on the appropriateness of the educational environment. With the growing awareness of problems relating to the negative influences of labelling, this is a serious consideration which must be examined closely.

2. The teacher has little input at the psychological assessment level. The educational component in the assessment process is greatly outweighed by the psychological component, and is often prepared by individuals having clinical experience which might not directly be related to educational programming. The concern must therefore be expressed that close scrutiny be given to the validity, reliability and educational significance of the tests in use, and especially to the manner in which the results of these tests are interpreted.

Summary and Conclusions

Definitional Problems in Learning Disabilities

The majority of the presently used definitions of Learning Disability, including that of the Ministry of Education, focus on the present functioning of a learner and particularly on a discrepancy between a pupil's academic achievement and some estimate of his or her potential. Although the earlier emphasis on etiological diagnosis has been reduced in importance, the majority of the assessment procedures in use are of the medical-clinical sort which purport to identify some causal factor or basic disorder in a learner's cognitive process. Usually also, these assessment procedures involve only analysis of present functioning without concern for either history or prognosis. In general, all existing definitions would be judged inadequate by the criteria of Cromwell et al (1975), since prognosis, either with or without a defined intervention, is not included in the construct.

A full or valid definition would have to include either a statement about an individual's predicted academic process without treatment, and/or a statement about the outcome to be expected following a specific treatment. Despite the extensive interest in and concern about learning disabilities over the last decade, we do not have the longitudinal data to make those statements. Beyond the simple observation that children who are doing poorly in school are not likely to become good learners within the regular curriculum - an observation which applies to almost all individuals experiencing school failure, not only the group defined as learning disabled -

we have little prognostic data. As for predictions based on applications of a particular psychoeducational treatment, we have seen a progression of programs based on motor retraining, perceptual training, and, more recently, psycholinguistic functioning, each showing promise and each being discounted eventually.

In the absence of agreement on a definition or empirical evidence that "learning disability" exists as a clinical entity independent of the learning environment and without having a clearly defined specific educational procedure for this group, it is little wonder the current assessment procedures are varied and confused and that the field is rife with conflicting opinions as to what should be assessed. None of the definitions in use clearly indicate an identification procedure, and so practitioners draw on the multiple historical strands which have been interwoven in the field and attempt to assess some or all of: gross-motor functioning, perceptual-motor functioning, learning potential, general intelligence, psycholinguistic functioning, academic performance, and, sometimes, adaptive behaviour. In spite of research evidence which consistently indicates that clinical judgement based on multiple data sources is usually less reliable than quantitative prediction based on a limited set of inputs (Sawyer, 1966), many school boards and diagnostic units continue to assemble complex batteries of assessment procedures, while others attempt to develop educational prescriptions which will be effective with a wide range of learners, therefore eliminating the need for individual assessment.

Validity of Assessment Procedures

A second major problem is the lack of substantial evidence for the validity of the common instruments in use with children suspected of having learning disabilities.

The focus of traditional psychoeducational diagnostic procedures has been on the determination of etiology--the basic cause of a learning disability. Toward this goal tests were selected and used without a firm theoretical and empirical base and diagnosticians have not always been careful enough in selecting and using tests for the purposes they were designed to serve or were capable of serving. Formal tests were taken over enthusiastically by educators as though important clues to educational programming resided in the test results. Mann (1971) raises serious theoretical questions about the assumptions underlying such testing. The methods of testing for learning disabilities also have been roundly criticized in an extremely detailed assessment of the field by Coles (1978), whose exacting review of the literature and evaluation of the techniques of diagnosis casts strong doubt on long-accepted tests of learning disability. Coles discusses the reliability and validity of such tests as the Bender Visual-Motor Gestalt Test, the Frostig Developmental Test of Visual Perception, the Wepman Auditory Discrimination test, Wechsler Intelligence Scale, etc., and cites numerous research reports regarding their usefulness. In keeping with the findings reported for each of these tests in the previous sections of this document, Coles concludes that there is very little empirical support for the tests and wonders why they are still being used for the purpose of drawing conclusion about learning disability and labelling students. He concludes his review by stating, "there is little question that eventually the tests reviewed here will be discarded..." (p.355).

At the moment, there is no one valid psychological test of a learning disability. Furthermore, the combination of a number of tests, contrary to conventional diagnostic thinking, does little to improve diagnostic efficiency. While many tests claim psychoeducational utility in diagnosing learning disability, all of these have such speculative psychometric properties that the experienced practitioner should be reluctant to employ them.

Toward an Alternative Look at Assessment

Perhaps the whole process of assessment needs to be reconsidered. Let us consider first the purpose for administering tests of any kind. In the education system, the major concern of any assessment procedure should be to provide relevant information which will assist the classroom teacher to facilitate student learning. In the psychoeducational tradition, the focus has tended to be upon the cause of disability and on remediating the condition in keeping with what has been termed the "medical model" (Mercer, 1979). In this approach, assessment emphasizes the deficits inherent in the child, sometimes giving little consideration to the educational, home and classroom context from which the child is drawn, or the effect that this disability may have had upon the child's overall educational endeavours. Regulation 704, under the Ontario Education Act, which governs the procedures for assessment, tends to reflect this approach. The regulation calls for each board to appoint a Special Education Program Placement and Review Committee to consider each child referred for special class placement. This committee is required to recommend placement, but is not required to provide educational direction such as suggesting teaching objectives or programs for the child. While the regulation does not explicitly prohibit committees from developing programming recommendations in addition to deciding placement, it does not encourage them to do so. It is therefore not surprising that few boards in Ontario are currently using their committees for recommending educational programs (Keeton, 1979). If a placement decision is the only goal of the assessment process, the medical or clinical diagnostic approach may appear to work, since it provides a label for the child which can be used to group him with similarly labelled children. The negative potential for such labelling has been the subject of much debate in the field (Hobbs, 1975). If the goal of assessment is to recommend effective instruction for the child, this clinical approach is clearly inadequate, since it fails to take account of the child's educational potential or provide for an effective intervention.

The composition of the Program Placement and Review Committee, suggested in Regulation 704 (Ministry of Education, 1979), further contributes to the lack of an educational focus in the assessment process. The committee must consist of at least three members, one of whom must be a supervisory officer. The involvement of the child's teachers and special education consultants is not specified. Indeed, under the now obsolete Regulation 191, educators were explicitly absent from the special education admissions committee, a policy which many Ontario boards continue to follow. In these boards, the committee consists of a superintendent, a principal and a psychologist or psychometrist (Keeton, 1979). In the current Regulation 704, the committee is charged with reaching a decision about placement. To do this, the board is required to supply information from the child's health and psychological assessment. A profile of the child's current educational standing or a report from his or her classroom teacher is still not required.

To a large degree, the medical and clinical orientation of assessment is preserved in the spirit of Regulation 704, in which the child's physiological and psychological deficits are focussed, and educational assessment and programming are likely to be the responsibility of the child's teachers only after a special class placement is made.

Assessment should ultimately be a team process: the sum of judgements and assessment by a group of educational and other professionals. Salvia & Ysseldyke (1978) consider the importance of professional judgement in assessment in contrast to the practice of interpreting a child's deficiencies from standardized test scores. They note:

> Diagnosticians seek out other professionals to complement their own skills and background. Thus referring a student to various specialists (audiologists, ophthamologists, reading teachers and so on) is a common and desirable practice in assessment. Judgements by teachers, counsellors, psychologists and practically any other school employees may be useful in particular circumstances. Expertise in making judgements is often a function of familiarity with the student being assessed. Teachers regularly express professional judgements... Judgements represent both the best and the worst of assessment data. Judgements made by conscientious, capable and objective individuals can be an invaluable aid in the assessment process. Inaccurate, biased, subjective judgements can be misleading at best and harmful at worst. Finally, all assessment ultimately require judgement. (p. 9, 10)

The point, as Salvia & Ysseldyke see it, is that the quality of psychoeducational assessment does not depend solely on the validity and accuracy of a standardized test instrument. Tests are only tools in the hands of professionals. The quality of assessment depends upon the competence of the professionals in evaluating test protocols to forecast the child's educational potential. Such potential is not predictable by identifying problems inherent in the child. It requires drawing from a broader educational context, from other professionals familiar with various aspects of the child's life and schooling. This context may require the involvement of medical personnel and other health professionals, the child's teachers, his parents and other caretakers, and frequently the child himself. After arriving at judgements and predictions about the child's educational potential, the child's progress should be closely monitored to check on the accuracy of the judgements and to make any adjustments to the educational provisions. Assessment is a human responsibility and not the product of tests. It is often an uncomfortable responsibility, and not easily undertaken. Its accuracy will depend on the cooperation and objectivity of a team of professionals, and upon the speed with which it can be responsive to the changing needs of the child.

Such a process requires carefully designed procedures[1] implemented on a board-wide basis which provide for the involvement of professionals as an assessment team. Since the team meets to consider cases, and to monitor progress, it is often most efficiently conducted on site in a family of schools. Teachers and parents can then be fully involved in deliberations and decision-making, and can report on children's progress according to the criteria specified by the team. Psychologists, psychometrists and health professionals act as resources to the educational assessment team, contributing test protocols and other information required to assist the team in recommending educational objectives and programs, and the appropriate placements for providing these. The results of standardized tests can then be used with the discretion they deserve, tempered with information gathered from other sources and with the professional opinions of the educational needs of the child.

---

[1] For excellent resources, see Council for Exceptional Children, Special Education Administrative Policies Manual (Reston, Virginia: CEC, 1977), and A.L. Pasanello, and C.B. Volkmor, Coming Back...or never leaving (Columbus, Ohio: Charles E. Merrill, 1977).

Final Note

Throughout this paper we have been concerned with a variety of technical measurement issues, in line with our charge to examine the validity of the assessment instruments and procedures used in the identification and placement of children with learning disabilities in this Province. Our conclusion, that there are no objective, valid measures which can be used in isolation for this purpose, will come as no surprise to those working in the special education field. In this area there is no way to avoid the realization that we are engaged in a process of weighing varied and sometimes discrepant information and making the best human judgement possible. We cannot, in conscience, hide this fact behind fixed cutting scores, general or specific tests or referrals to other professionals such as psychologists or medical doctors.

It is important, therefore, in attempting to identify and plan for those children whose school performance indicates the need for special programming, that as wide a range of types of information and expertise as possible be included. Some important data sources are:

- teachers, skilled in individualization of instruction, who can contribute refined educational assessment and suggestions for programming.
- parents, who can provide a view of a learner's adaptive capabilities in a variety of settings, and who can function as advocates for the child's best interests.
- psychologists, who can provide not only test scores and interpretations but also estimates of the degree

of confidence to be placed in such measures in a particular case.

- regular teachers, whose observation of students in depth provides complementary data to that of the individual assessors.
- administrators, who have an understanding of the possible educational alternatives in a particular school or school board.
- health professionals, who may have insight into nutritional or other health problems related to learning problems.

Because of the importance of the decisions we may be making regarding individual pupils, the absence of operationalized diagnostic procedures and the lack of validated measures, we must always realize that our placement decisions are simply the best we can do at the moment and resist the temptation to freeze them into permanent entity. Any of our decisions should be open to change and re-examination as the need arises, and any placement should be carefully monitored to ensure that the progress we hoped for actually occurs. Assessment and placement of a child with a learning disability is not the end of a process but merely the beginning.

Index to Tests Reviewed

Auditory Discrimination Test (Wepman) 70

(Beery-Buktenica) Developmental Test of Visual Motor Integration 68

Bender-Gestalt Test of Basic Skills 62

California Achievement Test 91, 92

Canadian Test of Basic Skills 80

Coloured Progressive Matrices (Raven's) 43

Denver Developmental Screening Test 38

Developmental Test of Visual Motor Integration (Beery-Buktenica) 68

Developmental Test of Visual Perception (Frostig) 66

Gates-MacGinitie Reading Tests 82

Illinois Test of Psycholinguistic Abilities, ITPA 51

McCarthy Scales of Children's Abilities 59

Metropolitan Achievement Test 92, 93

Peabody Individual Achievement Test 86, 90, 92

(Raven's) Coloured Progressive Matrices 43

(Slingerland) The Screening Test for Identifying Children with Specific Language Disability 64

Slosson Intelligence Test 49

Wechsler Intelligence Scale For Children -Revised (WISC-R) 29

Wepman Auditory Discrimination Test 70

Wide Range Achievement Test (WRAT) 76, 89

References

Allen, R. M. Factor analysis of the DTVP performance of EMR's. Perceptual and Motor Skills, 1968, 26, 257-258.

Anastasi, A. Psychological testing. New York: Macmillan Publishing Company, 1976.

Ashlock, P., & Thompson, G. The Ashlock tests of visual perception, revised. Paper presented at the International Scientific Conference of IFLD, 3rd., 1976, Montreal, Canada. (ERIC Document Reproduction Service No. 135 146.)

Ayres, A. Sensory integration and learning disorders. Los Angeles: Western Psychological Services, 1972.

Bannatyne, A. Diagnosis: A note on recategorization of the WISC Scaled Scores. Journal of Learning Disabilities, 1974, 7(5), 272-273.

Bannatyne, A. Peabody Individual Achievement Test. In O.K. Buros (Ed.), The eighth mental measurements yearbook. New Jersey: Gryphon Press, 1978.

Banks, C. Progressive Matrices. In O.K. Buros (Ed.), The fourth mental measurements yearbook. New Jersey: Gryphon Press, 1953.

Bateman, B. Learning disabilities--yesterday, today, and tomorrow. Exceptional Children, 1964, 31, 167-177.

Bateman, B. Reading and psycholinguistic processes of partially seeing children. C.E.C. Research Monographs, Series A, 1963, 5.

Bauman, E., & St. John, J. The clinical usefulness of some tests of visual perception. Psychology in the Schools, 1971, 8(3), 247-249.

Beatley, R. J. The relationship between cognitive tasks and school achievement. Dissertation Abstracts International, 1976, 36(11A), 397-399.

Beery, K. E. Preschool prediction and prevention of learning disabilities. Marin County Superintendent of Schools Office, March 1967. (ERIC Document Reproduction Service No. ED013 118.)

Belmont, I., & Belmont, L. Stability or change in reading achievement over time: Developmental and educational implications. Journal of Learning Disabilities, 1978, 11(2), 31-39.

Berg, P. C. Gates-MacGinitie Reading Tests: Readiness skills. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Birch, L. B. Canadian Test of Basic Skills. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Black, W. Neurological dysfunction and reading disorders. Journal of Learning Disabilities, 1973, 6 (5).

Blatt, B. Public policy and the education of children with special needs. Exceptional Children, 1972, 38, 537-543.

Bloom, B. S. Stability and change in human characteristics. New York: Wiley, 1964.

Boll, T. J. Correlation of WISC with motor speed and strength for brain-damaged and normal. The Journal of Psychology, 1971, 77, 162-172.

Book, R. M. Predicting reading failure: A screening battery for kindergarten children. Journal of Learning Disabilities, 1974, 7(1), 43-47.

Bortner, M. Raven Progressive Matrices. In O.K. Buros (Ed.), The sixth mental measurements yearbook. New Jersey: Gryphon Press, 1965.

Braithwaite, R. J. The Frostig Test and Training Programme--How valuable? Slow Learning Child, 1972, 19(2), 86-91.

Bray, N. M., & Estes, R. E. Achievement screening with learning disabled children. Paper presented at the Annual Meeting of the American Educational Research Association, 1975, Washington, D.C. (ERIC Document Reproduction Service No. ED112 576.)

Bray, N. M., & Estes, R. E. A comparison of the PIAT, CAT, and WRAT scores and teacher ratings for learning disabled children. Journal of Learning Disabilities, 1975, 8(8), 519-523.

Burke, C. L. Gates-MacGinitie Reading Tests. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Camp, B. W. Psychometric tests and learning in severely disabled readers. Journal of Learning Disabilities, 1973, 6(7), 53-58.

Carrol, J. Illinois Test of Psycholinguistic Abilities. In O.K. Buros (Ed), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Chissom, B. S. Marianne Frostig Developmental Test of Visual Perception, 3rd Edition. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Clark, C. M., & Dodd, B. E. Auditory factor in visual motor testing and training. Journal of Learning Disabilities, 1971, 4(10), 582-585.

Clements, S. Minimal brain dysfunction in children: Terminology and identification, phase one of a three-phase project. Washington, D.C.: United States Department of Health, Education and Welfare, 1966.

Coles, G. S. The learning disabilities test battery: Empirical and social issues. Harvard Educational Review, 1978, 48(3), 313-340.

Colligan, R. C. Concurrent validity of the Myklebust Pupil Rating Scale in a kindergarten population. *Journal of Learning Disabilities*, 1977, *10*, 317-320.

C.S.E. Elementary School Test Evaluation. Los Angeles: Center for the Study of Evaluation, University of California at Los Angeles, 1976.

Council for Exceptional Children, *Special Education Administrative Policies Manual*. Reston, Virginia.: CEC, 1977.

Council for Exceptional Children. Ensuring equality of educational opportunities in Ontario. A report of OCASE at the *Annual Provincial Meeting, C.E.C. Convention*, 1978, Ottawa, Ontario.

Coy, M. N. The Bender Visual-Motor Gestalt Test as a predictor of academic achievement. *Journal of Learning Disabilities,* 1974, *7*(5), 317-319.

Cromwell, R. L., Blashfield, R. K., & Strauss, J. S. Criteria for classification systems. In N. Hobbs (Ed.), *Issues in the classification of children* (Vol. 1). San Francisco: Jossey-Bass, 1975.

Deno, E. Screening Tests for Identifying Children with Specific Language Disability, Revised Edition. In O.K. Buros (Ed.), *The seventh mental measurements yearbook*. New Jersey: Gryphon Press, 1972.

DiCarlo, L. Auditory Discrimination Test. In O.K. Buros (Ed.), *The sixth mental measurements yearbook*. New Jersey: Gryphon Press, 1965.

Divoky, D. Is this screening test worth $1,319,638.50? *Learning*, 1977, *5*, 66-72.

Donah, C. E. H. Concurrent, predictive, and construct validation of the Slingerland Screening Tests for Identifying Children with Specific Language Disability (Doctoral dissertation, Kent State University, 1976). *University Microfilms International*, 1976 (University Microfilms International No. 77-7819).

Ducharme, D. J., & Toutant, R. L. *Evaluation of the Brant R.C.S.S. system approach to early identification of children with learning problems*. Ontario Institute for Studies in Education, Midwestern Centre, 1977.

Dunn, L. M., & Markwardt, F. C. *Peabody Individual Achievement Test Manual*. Circle Pines, Minn.: American Guidance Services, 1970.

Dykstra, R. Gates-MacGinitie Reading Tests: Readiness skills. In O.K. Buros (Ed.), *The seventh mental measurements yearbook*. New Jersey: Gryphon Press, 1972.

Edwards, R. P., Alley, G.R., & Snider, W. Academic achievement and minimal brain dysfunction. *Journal of Learning Disabilities*, 1971, *4*(3), 17-21.

Egland, B., Dinello, M., & Carr, D. The relationship of intelligence, visual-motor, psycholinguistic, and reading readiness skills with achievement. *Educational and Psychological Measurement*, 1970, *30*, 451-458.

Eller, W. Gates Primary Reading Tests. In O.K. Buros (Ed.), The sixth mental measurements yearbook. New Jersey: Gryphon Press, 1965.

Farrald, C., & Schamber, G. Problem identification through observation. Academic Therapy, 1973, 9(1), 85-89.

Ferdinand, W. E., & Jacobson, S. Early identification of learning disabilities. Journal of Learning Disabilities, 1970, 3(11), 589-593.

Finkenbinder, R. L. A descriptive study of the Goldman-Fristoe-Woodcock Test of Auditory Discrimination and selected reading variables with primary school children. Journal of Special Education, 1973, 7(2), 125-131.

Ford, J. New directions in special education. Journal of School Psychology, 1971, 9, 147-153.

Foster, G., Schmidt, C., & Sabatino, D. Teacher expectancies and the label "Learning Disabilities". Journal of Learning Disabilities, 1976, 9(2), 111-114.

Frankenburg, W. K. The Revised Denver Developmental Screening Test. Journal of Pediatrics, 1971, 79(6), 988-995.

Frankenburg, W. K., & Dodds, J. B. The reliability and stability of the Denver Developmental Screening Test. Child Development, 1971, 42 1315-1325.

Frankenburg, W. K., & Dodds, J. B. Validity of the Denver Developmental Screening Test. Child Development, 1971, 42, 475-485.

Frankenburg, W. K., & Dodds, J.B. Denver Developmental Screening Test, Reference Manual (Revised ed. 1975). Ladoca Project and Publishing Company, 1975.

French, J. L. Peabody Individual Achievement. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Frostig, M., & Horne, D. The developmental program in visual perception: Advanced pictures and patterns. Chicago: Follett, 1966.

Gallagher, J. Children with developmental imbalances: A psychoeducational definition. In W. Cruickshank (Ed.), The teacher of brain injured children: A discussion of the bases for competency. Syracuse, N.Y.: Syracuse University Press, 1966.

Gates, A., & MacGinitie, W. H. The Gates-MacGinitie Reading Tests. New York: Teacher's College Press, Columbia University, 1964.

Gearheart, B. Learning disabilities: Educational strategies. St. Louis, Mo.: C. V. Mosby, 1973.

Gershman, J., & Kershaw, J. The effects of language background and socio-economic status on screening procedures for the early identification of learning problems. Board of Education for the City of Toronto, February, 1977, Research Department, Report No. 147.

Gomez, E. H. Minimal cerebral dysfunction (maximum neurological confusion). Clinical Pediatrics, 1976, 6, 589-591.

Hagin, R. A., Silver, A. A., & Corwin, C. G. Clinical-diagnostic use of the WPPSI in predicting learning disabilities in Grade 1. Journal of Special Education, 1971, 5(3), 221-232.

Hallahan, D., & Cruickshank, W. Psychoeducational foundations of learning disabilities. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Hammill, D. D. Learning disabilities: A problem in definition. PRISE Reporter. Resource and Information Center for Special Education, Harrisburg, Pennsylvania, 1972, 1(4), 7-10.

Hammill, D., & Larsen, S. The relationship of selected auditory perceptual skills and reading ability. Journal of Learning Disabilities, 1974, 7, 429-436.

Haring, N. G., & Ridgway, R. W. Early identification of children with learning disabilities. Exceptional Children, 1967, 33(6), 387-395.

Harmer, W. R., & Wulliams, F. The Wide Range Achievement Test and the Peabody Individual Achievement Test: A comparative study. Journal of Learning Disabilities, 1978, 11(10), 65-68.

Harth, R., & Justen, J.E. The validity of the ITPA Visual Closure Subtest. Academic Therapy, 1977, 12, 261-265.

Hase, H. D. Predicting performance in the first grade with the First Grade Screening Test. Psychology in the Schools, 1977, 14(4), 407-412.

Hawthorne, L. W., & Larsen, S. C. The predictive validity and reliability of the Basic School Skills Inventory. Journal of Learning Disabilities, 1977, 10(1), 44-50.

Himelstein, P. Slosson Intelligence Test. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Hirshoren, A. A comparison of the predictive validity of the Revised Stanford-Binet Intelligence Scale and the Illinois Test of Psycholinguistic Ability. Exceptional Children, 1969, 25, 517-521.

Hobbs, N. (Ed.) Issues in the classification of children. San Francisco.: Jossey-Bass, 1975.

Hullsman, J. R. The WISC subtest syndrome for disabled readers. Perceptual and Motor Skills, 1970, 30, 535-550.

Hunt, J. V. Slosson Intelligence Test. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Hunt, J. V. McCarthy Scales of Children's Abilities. In O.K. Buros (Ed.), The eighth mental measurements yearbook. New Jersey: Gryphon Press, 1978.

Jansky, J. J. Early prediction of reading problems. Bulletin of the Orton Society, 1973, 23, 78-89.

Johnson, D., & Myklebust, H. Learning disabilities: Educational principles and practices. New York: Grune & Stratton, 1967.

Johnson, S. W., & Morasky, R. L. Learning disabilities. New Jersey: Allyn & Bacon Inc., 1977.

Kapelis, L. Early identification of reading failure: A comparison of two screening tests and teacher forecasts. Journal of Learning Disabilities, 1975, 8(10), 39-42.

Kass, C., & Myklebust, H. Learning disability: An education definition. Journal of Learning Disabilities, 1969, 2, 377-379.

Kass, C. E., & Macione, J. R. Psycholinguistic disabilities of children with reading problems. Exceptional Children, 1966, 32, 533-539.

Kaufman, A. A new approach to the interpretation of test scatter on the WISC-R. Journal of Learning Disabilities, 1976, 9(3).

Kaufman, A. Verbal-Performance IQ discrepancies on the WISC-R. Journal of Consulting and Clinical Psychology, 1976, 44(5), 739-744.

Kaufman, A. S., & Kaufman, N. L. Research on the McCarthy Scales and its implications for assessment. Journal of Learning Disabilities, 1977, 10(5), 284-291.

Kaufman, N. L. Use of the WISC-R with a learning disabled population: Some diagnostic implications. Psychology in the Schools, 1976, 13(4), 381-386.

Keeton, A., Special Education: A right or a privilege? Interchange, 1979, (in press).

Kenny, T. J., & Clemmens, R. Medical and psychological correlates in children with learning disabilities. Journal of Pediatrics, 1971, 78(2), 273-277.

Keogh, B. K. Early ID: Selective perception or perceptive selection? Academic Therapy, 1977, 12(3), 267-274.

Keogh, B., & Hall, R. WISC subtest patterns of educationally handicapped and educable mentally retarded pupils. Psychology in the Schools, 1974, 11, 296-300.

Kephart, N. C. The slow learner in the classroom. Columbus, Ohio: Merrill, 1971.

Kephart, N.C. Marianne Frostig Developmental Test of Visual Perception. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Kirk, S. Educating exceptional children (1st Ed.). Boston: Houghton Mifflin, 1962.

Kirk, S. Educating exceptional children (2nd Ed.). Boston: Houghton Mifflin, 1972.

Kirk, S., & Bateman, B. Diagnosis and remediation of learning disabilities. *Exceptional Children*, 1962, *29*(2), 73-78.

Kirk, S., & Kirk, W. *Psycholinguistic learning disabilities: Diagnosis and remediation*. Urbana, Ill.: University of Illinois Press, 1971.

Kirk, S., & McCarthy, J. *Examiners' manual: Illinois Test of Psycholinguistic Abilities*. Urbana, Ill.: University of Illinois Press, 1969.

Kirk, S., McCarthy, J., & Kirk, W. *Illinois Test of Psycholinguistic Abilities*. Urbana, Ill.: University of Illinois Press, 1968.

Kirk, S. A., & Elkins, J. Identifying developmental discrepancies at the preschool level. *Journal of Learning Disabilities*, 1975, *8*(7), 417-419.

Kirschenbaum, D. S. The effectiveness of a mass screening procedure in an early intervention program. *Psychology in the Schools*, 1977, *14*, 400-406.

Kitay, P. M. Bender Gestalt. In O.K. Buros (Ed.), *The seventh mental measurements yearbook*. New Jersey: Gryphon Press, 1972.

Koenke, K. A comparison of three auditory discrimination-perception tests. *Academic Therapy*, 1978, *13*, 463-468.

Koppitz, E. M. Bender Gestalt Test, Visual Aural Digit Span Test and reading achievement. *Journal of Learning Disabilities*, 1975, *8*(3), 154-158.

Kratochwill, T. R., & Demuth, D. M. An examination of the predictive validity of the Keymath Diagnostic Arithmetic Test and the Wide Range Achievement Test in exceptional children. *Psychology in the Schools*, 1976, *13*(4), 404-406.

Landrus, G. *Toronto Early Identification and Development Program*. Ministry of Education, Ontario, NO. ONOO 321.

Larsen, S. C. The use of selected perceptual tests in differentiating between normal and learning disabled children. *Journal of Learning Disabilities*, 1976, *9*(2), 85-90.

Lerner, J. *Children with learning disabilities: Theories, diagnosis, and teaching strategies*. Boston: Houghton Mifflin, 1971.

Lester, M. Marianne Frostig Developmental Test of Visual Perception, 3rd Edition. In O.K. Buros (Ed.), *The seventh mental measurements yearbook*. New Jersey: Gryphon Press, 1972.

Leton, D. A. Developmental Test of Visual-Motor Integration. In O.K. Buros (Ed.), *The eighth mental measurements yearbook*. New Jersey: Gryphon Press, 1978.

Lindquist, E. F., & Hieronymus, A. N., adapted by E. King. *Canadian Tests of Basic Skills*. Toronto: Thomas Nelson & Sons (Canada) Ltd., 1976.

Lowrance, D. S. Psycho-educational assessment of the learning disabled child (Doctoral dissertation, University of Southern Mississippi, 1977). *University Microfilms International*, 1977. (University Microfilms International No. 77-22, 878.)

Lyman, H. B. Peabody Individual Achievement Test. In O.K. Buros (Ed.), *The seventh mental measurements yearbook*. New Jersey: Gryphon Press, 1972.

Mann, L.M. Psychometric phrenology and the new faculty psychology: The case against ability testing and training. *Journal of Special Education*, 1971, 5, No. 1.

Mardell, C., & Goldenberg, D. For prekindergarten screening information: DIAL. *Journal of Learning Disabilities*, 1975, *8*(3), 140-147.

Marshall, N., Hopps, D., & McKinnon, G. Language abilities of children with differing socioeconomic backgrounds. *Journal of Learning Disabilities*, 1970, *3*(6), 334.

Mauser, A. J. *Assessing the learning disabled: Selected instruments.* San Rafael, California: Academic Therapy Publications, 1976.

McCarthy, J., & Kirk, S. *The construction, standardization and statistical characteristics of the Illinois Test of Psycholinguistic Abilities.* Urbana, Illinois: University of Illinois Press, 1963.

McCarthy, J. J., & Larsen, S. C. The validity of perceptual tests: The debate continues. Rebuttal. *Journal of Learning Disabilities*, 1976. *9*(6), 332-337.

McCarthy, J. J., & McCarthy, J. F. *Learning Disabilities*. Boston: Allyn & Bacon, 1969.

McIntosh, D., & Dunn, L. Children with major specific learning disabilities. In L. Dunn (Ed.), *Exceptional children in the schools*. New York: Holt, Rinehart & Winston, 1973.

Mercer, C., Forgone, C., & Wolking, W. D. Definitions of learning disabilities used in the United States. *Journal of Learning Disabilities*, 1976, *9*(6), 376-386.

Mercer, J.R. *System of Multicultural Pluralistic Assessment*. New York: The Psychological Corporation, 1979.

Mercer, J. (as reported in Foster, et al., 1976). *The labeling process*. Paper presented at the Joseph P. Kennedy, Jr., Foundation International Symposium on Human Rights, Washington, D. C., 1971.

Miller, M. WISC Subtest patterns as discriminators of perceptual disabilities. *Journal of Learning Disabilities*, 1978, *11*(7).

Ministry of Education, Memorandum 1978-79: 15. Re: Early identification of children's learning needs. December 27, 1978.

Ministry of Education, Memorandum 1978-79: 14. Re: The education of students with learning disabilities: Statement to School Boards. December 21, 1978.

Mittler, P., & Ward, J. The use of the ITPA on British 4 year old children: A normative and factorial study. British Journal of Educational Psychology, 1970, 40, 43-54.

Moriarty, A. E. Denver Developmental Screening Test. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Mueller, M. W. Prediction of achievement of educable mentally retarded children. American Journal of Mental Deficiency, 1969, 73, 590-596.

Munns, E. F. The development of a teachers' observation scale for the identification of children with learning disabilities (Doctoral dissertation, York University, 1971). National Library of Canada, 1971. (ERIC No. EC093469)

Myers, P., & Hammill, D. D. Methods for learning disorders. New York: John Wiley, 1969.

Myklebust, H. R. The pupil rating scale: Screening for learning disabilities. New York: Grune & Stratton, Inc., 1971.

National Advisory Committee on Handicapped Children. First Annual Report. Washington, D. C.: U.S. Department of Health, Education and Welfare, 1969.

Newcomer, P. Construct validity of the Illinois Test of Psycholinguistic Abilities. Journal of Learning Disabilities, 1975, 8(4), 220-231.

Newcomer, P., & Hammill, D. A short form of the Revised Illinois Test of Psycholinguistic Abilities. Journal of Learning Disabilities, 1974, 1(9), 570-572.

Newcomer, P., & Hammill, D. ITPA and academic achievement: a survey. Reading Teacher, 1975, 28.

Newcomer, P. L., & Hammill, D. D. Psycholinguistics in the schools. Ohio: Charles E. Merrill Publishing Co., 1976.

Noonan, J. T. Specific predictors of learning disabilities. Academic Therapy, 1977, 12(4), 471-477.

O'Bryan, K. G. The early identification projects. Ministry of Education, Ontario, No. ON00092.

O'Bryan, K. G. The Windsor early identification project. Ministry of Education, Ontario, 1976, No. ON00565.

Okada, D. The effects of perceptual and perceptual motor training of the visual perception, auditory perception, and language performance of institutional educable mental retardates (Doctoral dissertation, New York University, 1969). Dissertation Abstracts International, 1969, 30, 2857-A. (University Microfilms No. 70-740.)

Orlando, C., & Lynch, J. Learning disabilities or educational casualties? Where do we go from here? Elementary School Journal, 1974, 74, 461-467.

Osgood, C. A Behavioristic Analysis In Contemporary Approach To Cognition. Cambridge, Mass.: Harvard University Press, 1977.

Paraskevopoulos, J.N., & Kirk, S.A. The development and psychometric characteristics of the revised Illinois Test of Psycholinguistic Abilities. Illinois: University of Illinois Press, 1969.

Polatajko, H., & Reich, C. The Medvedeff Program: An Analysis and Critique. O.I.S.E., Department of Special Education, January 1978.

Powell, W.R. Gates-MacGinitie Reading Tests. In O.K. Buros (Ed.), The seventh mental measurement yearbook. New Jersey: Gryphon Press, 1972, 1082-1083.

Proger, B.B. Peabody Individual Achievement Test. In O.K. Buros (Ed.), The eight mental measurements yearbook. New Jersey: Gryphon Press, 1978, 75-77.

Ranck, S.A. A Factor Analysis of Fine Perceptual Motor Tests of First Grade Children (Doctoral dissertation, Fordham University, 1976). Xeron University Microfilms, 1976 (Xeron University Microfilms No. 76-17, 914).

Raven, J.C. Guide to using the Coloured Progressive Matrices Sets A, A6, B (Revised Order 1956). London: H.K. Lewis & Company, 1965.

Rice, J.A. Developmental Test of Visual-Motor Integration. In O.K. Buros (Ed.), The eight mental measurements yearbook. New Jersey: Gryphon Press, 1978, 1400-1401.

Rourke, B.P. The Neuropsychological Abilities of Normal and Retarded Readers. Ministry of Education, Ontario ON00166.

Russell, E.W. WAIS Factor Analysis with Brain-Damaged Subjects Using Criterion Measures. Journal of Consulting and Clinical Psychology, Aug. 1972, 39, No. 1, 133-139.

Sabatino, D.A. Learning Disabilities Handbook: A Technical Guide to Program Development. Illinois: Northern Illinois University Press, 1976.

Sabatino, D.A., Abbott, J.C., & Becker, J.T. What Does the Frostig DTVP Measure? Exceptional Children, Mar. 1974, 40, No. 6, 453-54.

Sattler, J.M. Assessment of Children's Intelligence. Toronto: W.B. Saunders Company, 1974.

Sattler, J.M. McCarthy Scales of Children's Abilities. In O.K. Buros (Ed.), The eight mental measurements yearbook. New Jersey: Gryphon Press, 1978, 311-313.

Sawyer, J. Measurement and prediction, clinical and statistical. Psychological Bulletin, 1966, 66, 178-200.

Sears, C. A comparison of the basic language concepts and psycholinguistic abilities of second grade boys who demonstrated average and below average levels of reading achievement (Doctoral dissertation, Colorado State College, 1969). *Dissertation Abstracts International*, 1969, *30* 1758-A (University Microfilms No. 69-19.233).

Severson, R. A. Early detection of children with potential learning disabilities: A seven-year effort. *Proceedings of the 80th Annual Convention of the American Psychological Association*, 1972, *7*, 561-562.

Silverstein, A. B. McCarthy Scales of Children's Abilities. In O.K. Buros (Ed.), *The eighth mental measurements yearbook*. New Jersey: Gryphon Press, 1978.

Simensen, R. J., & Sutherland, J. Psychological assessment of brain damage: The Wechsler Scales. *Academic Therapy*, 1974, *10*(1), 69-81.

Sitlington, P. L. *Validity of the Peabody Individual Achievement Test for educable mentally retarded adolescents*. Unpublished Master's thesis, University of Hawaii, 1970.

Smith, C. The relationship of reading methods and reading achievement of ITPA sensory modalities (Doctoral dissertation, University of Georgia, 1969). *Dissertation Abstracts International*, 1970, *30* 2916-A. (University Microfilms No. 70-1194)

Smith, P. A., & Marx, R. W. Some cautions on the use of the Frostig test: A factor analytic study. *Journal of Learning Disabilities*, 1972, *5*(6), 357-362.

Soethe, J. W. Concurrent validity of the Peabody Individual Achievement Test. *Journal of Learning Disabilities*, 1972, *5*(9), 560-562.

Spalding, N. V. Research with QNST: A review. Paper presented at the *International Scientific Conference of IFLD*, 3rd, 1976, Montreal Canada. (ERIC Document Reproduction Service No. ED 135 153)

Spalding, N. V. S. The validition of a neurological screening test for learning disabilities (Doctoral dissertation, University of California at Berkeley, 1972). *University Microfilms International*, 1972 (University Microfilms International No. 77-7806).

Sterling, H. M., & Sterling, P. J. Experiences with the QNST. *Academic Therapy*, 1977, *12*, 339-342.

Sterritt, G. M., Martin, V., & Rudnick, M. Auditory-visual and temporal spatial integration as determinants of test difficulty. *Psychonomic Science*, 1971, *23*, 289-291.

Sterritt, G. M., & Rudnick, M. Reply to Birch and Belmont. *Perceptual and Motor Skills*, 1966, *23*, 662.

Strauss, A., & Kephart, N. *Psychopathology and education of the brain-injured child* (Vol. 2). New York: Grune & Stratton, 1955.

Swanson, M. S., & Jacobson, A. Evaluation of the S.I.T. for screening children with learning disabilities. Journal of Learning Disabilities, 1970, 3(6), 318-320.

Tauber, R. Identification of potential learning disabilities. Academic Therapy Quarterly, 1966-67, 2(2), 116-119.

Thorpe, H. S., & Werner, E. E. Developmental screening of preschool children: A critical review of inventories used in health and educational programs. Pediatrics, 1974, 53(3), 362-370.

Thweatt, R. C., Obrzut, J. E., & Taylor, H. D. The development and validation of a soft-sign scoring system for the Bender Gestalt. Psychology in the Schools, 1972, 9(2), 170-174.

Traxler, A. E. Sequential studies of pupil achievement. In H.M. Robinson (Ed.), Sequential development of reading abilities. Supplementary Education Monographs, No. 9, Chicago: University of Chicago Press, 1960.

Uhl, N. P., & Nurss, J. R. Socio-economic level styles in solving reading-related tasks. Reading Research Quarterly, 1970, 5, 452-485.

Valett, R. Programming learning disabilities. Belmont, California: Fearon Publishers, Inc., 1969.

Vance, H. Analysis of cognitive abilities for learning disabled children. Psychology in the School, 1976, 13(4), 477-483.

Van Roekel, B. H. Gates-McGinitie Reading Tests. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Waksman, M., Silverman, H., & Weber, K. Assessing the learning potential of penitentiary inmates on application of Feuerstein's Learning Potential Assessment Device. Report ot the Educational and Training Division of the Correctional Service of Canada, Ottawa, Canada, 1979.

Wall, W. D. Progressive Matrices. In O.K. Buros (Ed.), The fourth mental measurements yearbook. New Jersey: Gryphon Press, 1953.

Washington, E. D., & Teska, J. A. Correlations between the Wide Range Achievement Test, the California Achievement Tests, the Stanford Binet, and the Illinois Test of Psycholinguistic Abilities. Psychological Reports, 1970, 26, 291-294.

Watson, P. C. Congruent validity of the Peabody Individual Achievement Test as measured by the California Achievement Test with educationally handicapped and educable mentally retarded children. Unpublished Master's thesis, University of California at Los Angeles, 1972.

Wayne, D. A. The learning disabled adolescent in Ontario: Overview of the problem; assessment, diagnosis and placement; remediation; recommendations. Sponsored by Ontario Educational Research Council, Ontario Secondary School Teachers' Federation, 1978.

Welch, M., & Cook, J. M. Diagnostic and identification implications of primary mental abilities tests for educationally deficient children. U.S. Department of Health, Education , and Welfare, National Institute of Education, Washington, D. C., 1970. (ERIC Document Reproduction Service No. ED 109 860.)

Wepman, J. M. Screening Tests for Identifying Children With Specific Language Disabilities, Revised Edition. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Werner, E. E. Denver Developmental Screening Test. In O.K. Buros (Ed.), The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Werner, H., & Strauss, A. Disorders of conceptual thinking in the brain-injured child. Journal of Nervous and Mental Disease, 1942, 96, 153-172.

Westby, G. Progressive Matrices. In O.K. Buros (Ed.), The fourth mental measurements yearbook. New Jersey: Gryphon Press, 1953.

Westinghouse Learning Corporation/Ohio University. The impact of Head Start on children's cognitive and affective development (Vols. 1 & 2). Springfield, Virginia: U.S. Department of Commerce Clearinghouse, 1969.

Wetter, J., & French, R. W. Comparison of the Peabody Individual Achievement Test and the Wide Range Achievement Test in a learning disability clinic. Psychology in the Schools, 1973, 10(3), 285-286.

Wiedl, K. H., & Carlson, J. S. The factorial structure of the Raven Coloured Progressive Matrices Test. Educational and Psychological Measurement, 1976, 36, 409-413.

Wilson, J. D., & Spangler, P. F. The Peabody Individual Achievement Test as a clinical tool. Journal of Learning Disabilities, 1974, 7(6), 384-387.

Wisland, M., & Many, W. A. A factorial study of the ITPA with children having above average intelligence. Education and Psychology Measurement, 1969, 29, 367-376.

Zach, L., & Kaufman, J. How adequate is the concept of perceptual deficit for education? Journal of Learning Disabilities, 1972, 5(6), 351-356.

Zbindn, W. R. Psycholinguistic and perceptual correlates of spelling in educable mentally handicapped children (Unpublished doctoral dissertation, University of Illinois, 1970). University Microfilms,